



QY	61	GTGATGAGACGTGTCCCACTGAGGTGCCCAACAGCAGAGTGTGAGCATGGGCTGAG	120
Db	61	GTGATGAGACGTGTCCCACTGAGGTGCCCAACAGCAGAGTGTGAGCATGGGCTGAG	120
QY	121	AGCTGGACGGCACAAAGGCTGGCAGAAATGGGCGCCTGGCTGATTCCTAGGCAGTT	180
Db	121	AGCTGGACGGCACAAAGGCTGGCAGAAATGGGCGCCTGGCTGATTCCTAGGCAGTT	180
QY	181	GCGGGCAGCAAGAGAGAGGCGCAGCTTCTGGAGCAGAGCCGAGAGCAGTGTCTG	240
Db	181	GCGGGCAGCAAGAGAGAGGCGCAGCTTCTGGAGCAGAGCCGAGAGCAGTGTCTG	240
QY	241	GAGTGCCTGAAGGGCCCCGTAGGCCCTACCGCCCTGGCCCACTATGGTCCAGAGGCTGTG	300
Db	241	GAGTGCCTGAAGGGCCCCGTAGGCCCTACCGCCCTGGCCCACTATGGTCCAGAGGCTGTG	300
QY	301	GGTAGCCCGCTGTGCGGCACCGGAAAGCCCAAGCTTGTGTGCTCAACCTGTAAACCTT	360
Db	301	GGTAGCCCGCTGTGCGGCACCGGAAAGCCCAAGCTTGTGTGCTCAACCTGTAAACCTT	360
QY	361	TGSCCTTGGAGGTGTGTTTGGCCGCAAGGCATCACCTATATGCGCGCTCTGTGCTGGAAGT	420
Db	361	TGSCCTTGGAGGTGTGTTTGGCCGCAAGGCATCACCTATATGCGCGCTCTGTGCTGGAAGT	420
QY	421	GGGGGTAGAGGAGAGTTTATGACCATGGTGTGGGCATTTGGTCAGTGTGGCCCTGGT	480
Db	421	GGGGGTAGAGGAGAGTTTATGACCATGGTGTGGGCATTTGGTCAGTGTGGCCCTGGT	480
QY	481	CTGTGTCGGCTCTAGGCTCAGCAGTACACCTATGCGCGCTCTGTGCTGGAAGT	540
Db	481	CTGTGTCGGCTCTAGGCTCAGCAGTACACCTATGCGCGCTCTGTGCTGGAAGT	540
QY	541	GCCTTTTCATCTGGGCACTGTCTTGGGCATCTGCTGAGCCCTTCTTCATCCCAAGGGC	600
Db	541	GCCTTTTCATCTGGGCACTGTCTTGGGCATCTGCTGAGCCCTTCTTCATCCCAAGGGC	600
QY	601	CGSGTGGCTAGCAGGCTGTCTGTGCGGATCCAGGCGCTTGGAGCTGGGCACTGTCTAT	660
Db	601	CGSGTGGCTAGCAGGCTGTCTGTGCGGATCCAGGCGCTTGGAGCTGGGCACTGTCTAT	660
QY	661	CTTGGGCGTGGGCTGTCTGTGGGCATCTGCTGAGCCCTTCTTCATCCCAAGGGC	720
Db	661	CTTGGGCGTGGGCTGTCTGTGGGCATCTGCTGAGCCCTTCTTCATCCCAAGGGC	720
QY	721	GCCTCTGACCTCTCCGGGACCCGAGCAGTGTGCGGAGGCTACTGTCTATGCTT	780
Db	721	GCCTCTGACCTCTCCGGGACCCGAGCAGTGTGCGGAGGCTACTGTCTATGCTT	780
QY	781	CATGATCAGTCTTGGGGGTGCTGGGCTACCTCTGCTGCGCATTTGACTGGGACACAG	840
Db	781	CATGATCAGTCTTGGGGGTGCTGGGCTACCTCTGCTGCGCATTTGACTGGGACACAG	840
QY	841	TGGCTTGGCCCTACCTGGGCACCCAGAGAGTGCCTCTTGGGCTGCTCACCTCAT	900
Db	841	TGGCTTGGCCCTACCTGGGCACCCAGAGAGTGCCTCTTGGGCTGCTCACCTCAT	900
QY	901	CTTCTCTCAGCTGTAGCAGCCACACTGTGTGTGGCTGAGGAGGAGCGCTGGGCCCCAC	960
Db	901	CTTCTCTCAGCTGTAGCAGCCACACTGTGTGTGGCTGAGGAGGAGCGCTGGGCCCCAC	960
QY	961	CGAGCCAGCAGAGGCGTGTGGGCCCCCTTCTGTGCGCCCACTGCTTCCATGCGGGG	1020
Db	961	CGAGCCAGCAGAGGCGTGTGGGCCCCCTTCTGTGCGCCCACTGCTTCCATGCGGGG	1020
QY	1021	CCGCTTGGCTTCCGGAACCTGGGCGCCCTGCTTCCCGGCTGACACAGCTGTGCTGCG	1080
Db	1021	CCGCTTGGCTTCCGGAACCTGGGCGCCCTGCTTCCCGGCTGACACAGCTGTGCTGCG	1080
QY	1081	CATGCCCGGACCCCTGGGCGGCTTCTTGTGGCTGTGAGCTGTGCAGTGTGCACATCAT	1140
Db	1081	CATGCCCGGACCCCTGGGCGGCTTCTTGTGGCTGTGAGCTGTGCAGTGTGCACATCAT	1140

QY 1249 GATGGGAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGTCTAT 1308  
DB 2811 GATGGGAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGTCTAT 2870  
QY 1309 GGACCGGCTGGTGCACGAGTTCGGCACTCGACAGTCTATTGGGCGAGTGGGAGCTTT 1368  
DB 2871 GGACCGGCTGGTGCACGAGTTCGGCACTCGACAGTCTATTGGGCGAGTGGGAGCTTT 2930  
QY 1369 CCTGTGGCTGCCGTGGCCACATGCTGTCCACAGTGTGGCGGTGCACAGCTTCAGC 1428  
DB 2931 CCTGTGGCTGCCGTGGCCACATGCTGTCCACAGTGTGGCGGTGCACAGCTTCAGC 2990  
QY 1429 CGCCCTACCGGGTTCACCTTCTACGCCCTGCAGATCTCTGCCCTACACACTGGCCCTCC 1488  
DB 2991 CGCCCTACCGGGTTCACCTTCTACGCCCTGCAGATCTCTGCCCTACACACTGGCCCTCC 3050  
QY 1489 CTACACCGGGAGACGAGTGTCTGTGCCCAATACCCAGGGGACACTGGAGTGTCTAG 1548  
DB 3051 CTACACCGGGAGACGAGTGTCTGTGCCCAATACCCAGGGGACACTGGAGTGTCTAG 3110  
QY 1549 CAGTAGGACAGCTGATGACAGCTTCTGTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCC 1608  
DB 3111 CAGTAGGACAGCTGATGACAGCTTCTGTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCC 3170  
QY 1609 TAATGACACGTGGGTGCTGGAGGACAGTGGCCCTGTCTCCACCTCCACCGCGCTCTGGGG 1668  
DB 3171 TAATGACACGTGGGTGCTGGAGGACAGTGGCCCTGTCTCCACCTCCACCGCGCTCTGGGG 3230  
QY 1669 GGCCCTCTGCTGTATGCTCTCGTAGCTGTGGTGGTGGTGGAGCCACCGAGGCCAGGCT 1728  
DB 3231 GGCCCTCTGCTGTATGCTCTCGTAGCTGTGGTGGTGGTGGAGCCACCGAGGCCAGGCT 3290  
QY 1729 GGTTCGGCGCGGGGATCTGCTGGACCTCCGATCCGATCCGATGCTTCTCTCTGTC 1788  
DB 3291 GGTTCGGCGCGGGGATCTGCTGGACCTCCGATCCGATGCTTCTCTCTGTC 3350  
QY 1789 CCAGGTGGCCCATCCCTGTTTATGGGTCCATGTCAGCTCAGCCAGTCTGTCACTGC 1848  
DB 3351 CCAGGTGGCCCATCCCTGTTTATGGGTCCATGTCAGCTCAGCCAGTCTGTCACTGC 3410  
QY 1849 CTATATGCTGTCCGCGAGGCTGGTCTGTGTGGTCCATGTCAGCTCAGCCAGGATG 1908  
DB 3411 CTATATGCTGTCCGCGAGGCTGGTCTGTGTGGTCCATGTCAGCTCAGCCAGGATG 3470  
QY 1909 ATTTGACAGAGCGACTTGGCCAAATACCTACGCTAGAAAATCTCCAGCACATGGGGTG 1968  
DB 3471 ATTTGACAGAGCGACTTGGCCAAATACCTACGCTAGAAAATCTCCAGCACATGGGGTG 3530  
QY 1969 GAGGGCTGCCTCACTGGGTCCAGCTCCCGCTCTCTGTTAGCCCATGGGGCTGCCGG 2028  
DB 3531 GAGGGCTGCCTCACTGGGTCCAGCTCCCGCTCTCTGTTAGCCCATGGGGCTGCCGG 3590  
QY 2029 CTGGCCGCGAGTTCTGTGTCTGCCAAGTATGTGGTCTCTGTGTCACACCTGTGCTG 2088  
DB 3591 CTGGCCGCGAGTTCTGTGTCTGCCAAGTATGTGGTCTCTGTGTCACACCTGTGCTG 3650  
QY 2089 CTGAGTGTGCTAGCTGCACAGCTGGGGCTGGGGCTCCCTCTCTCTCTCCCAAGTCTC 2148  
DB 3651 CTGAGTGTGCTAGCTGCACAGCTGGGGCTGGGGCTGGGGCTCCCTCTCTCTCTCCCAAGTCTC 3710  
QY 2149 TAGGGCTGCCTGACTGGAGGCTTCCAGGGGTTTCACTGTGGACTTATACAGGAGGC 2208  
DB 3711 TAGGGCTGCCTGACTGGAGGCTTCCAGGGGTTTCACTGTGGACTTATACAGGAGGC 3770  
QY 2209 CAGAAGGCTCCATGACTGGAAATGGGGGACTCTGCAGGTGGATTAACCGGCTCAGGG 2268  
DB 3771 CAGAAGGCTCCATGACTGGAAATGGGGGACTCTGCAGGTGGATTAACCGGCTCAGGG 3830  
QY 2269 TTAACAGCTAGCTCTCTAGTGTGACACACACTAGAGAGGGTTTGGGAGCTGAATAA 2328  
DB 3831 TTAACAGCTAGCTCTCTAGTGTGACACACACTAGAGAGGGTTTGGGAGCTGAATAA 3890

QY 2329 CTAGTCACTCCGTGGTTTCCCATCTCTAAGCCCTTAACTGCAGCTTCGTTTAAATGTAGCT 2388  
DB 3891 CTAGTCACTCCGTGGTTTCCCATCTCTAAGCCCTTAACTGCAGCTTCGTTTAAATGTAGCT 3950  
QY 2389 CTTCGATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATTTGAACATAATG--ACTTA 2446  
DB 3951 CTTCGATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATTTGAACATAATGAAGTTA 4010  
QY 2447 TTTGTAGGGAGAGTCTCTGAGGGGCAACACACAGAACAGGTCCTCCCTCAGGCCACAGC 2506  
DB 4011 TTTGTAGGGAGAGTCTCTGAGGGGCAACACACAGAACAGGTCCTCCCTCAGGCCACAGC 4070  
QY 2507 ACTGTCTTTTGTGCTATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTGGTCC 2566  
DB 4071 ACTGTCTTTTGTGCTATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTGGTCC 4130  
QY 2567 TTTCTGTTCCTATCAGAGACACAGGCAATTAATAATTTAACTTATTTTAAACAAGT 2626  
DB 4131 TTTCTGTTCCTATCAGAGACACAGGCAATTTAAATATTTTAACTTATTTTAAACAAGT 4190  
QY 2627 AGAAGGGAATCCATTGCTAGCTTTCTGTGTGTGTCTAATATTTGGTAGGGTGGGG 2686  
DB 4191 AGAAGGGAATCCATTGCTAGCTTTCTGTGTGTGTCTAATATTTGGTAGGGTGGGG 4250  
QY 2687 ATCCCAACAATCAGGTCCTCGATAGCTGGTCAATTTGGGCTGATCATTTGCCAGAATCT 2746  
DB 4251 ATCCCAACAATCAGGTCCTCGATAGCTGGTCAATTTGGGCTGATCATTTGCCAGAATCT 4310  
QY 2747 TCTTCTCTGGGCTGTGGCCCCCAAAATGCTTAACCCAGGACCTTGGAAATTTCTACTCA 2806  
DB 4311 TCTTCTCTGGGCTGTGGCCCCCAAAATGCTTAACCCAGGACCTTGGAAATTTCTACTCA 4370  
QY 2807 TCCCAAAATGAATTCCAAATGCTTTACCAAGGTTAGGTTGTGAAGGAGTACAGG 2866  
DB 4371 TCCCAAAATGAATTCCAAATGCTTTACCAAGGTTAGGTTGTGAAGGAGTACAGG 4430  
QY 2867 GTGGGGCTTCAGGCTCAACGGCTTCCCTAACCAACCCCTCTCTCTTTGGCCACGCTGGT 2926  
DB 4431 GTGGGGCTTCAGGCTCAACGGCTTCCCTAACCAACCCCTCTCTCTTTGGCCACGCTGGT 4490  
QY 2927 TCCCCCACTTCCACTCCCTCTACTCTCTAGGACTGGGCTGATGAAGGCACTGCCCA 2986  
DB 4491 TCCCCCACTTCCACTCCCTCTACTCTCTAGGACTGGGCTGATGAAGGCACTGCCCA 4550  
QY 2987 AAATTTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCCACGAGCTCCCAACCC 3046  
DB 4551 AAATTTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCCACGAGCTCCCAACCC 4610  
QY 3047 TGTGTTGGAGCTACTGCAGGACAGAGCAAAAGTGGGGTTTCCCAAGCCTTTTGTCCATC 3106  
DB 4611 TGTGTTGGAGCTACTGCAGGACAGAGCAAAAGTGGGGTTTCCCAAGCCTTTTGTCCATC 4670  
QY 3107 TCAGCCGCCAGAGTATATCTGTGTTGGGAACTCACACAGAACTCAGAGGACCCCCC 3166  
DB 4671 TCAGCCGCCAGAGTATATCTGTGTTGGGAACTCACACAGAACTCAGAGGACCCCCC 4730  
QY 3167 TGCTGTAGCTAAGGAGGCTTATCTCTCAGGGGGGTTTAAAGTCCGTTTGAATATG 3226  
DB 4731 TGCTGTAGCTAAGGAGGCTTATCTCTCAGGGGGGTTTAAAGTCCGTTTGAATATG 4790  
QY 3227 TCGTCTTATTTATTTAGCGGGTGAATATTTTACTGTAACTGAGCAATCAGAGTATAA 3286  
DB 4791 TCGTCTTATTTATTTAGCGGGTGAATATTTTACTGTAACTGAGCAATCAGAGTATAA 4850  
QY 3287 TGTGTTAGTGACAAAAATTAAGGCTTCTTATATGTTTAAAAA 3330  
DB 4851 TGTGTTAGTGACAAAAATTAAGGCTTCTTATATGTTTAAAAA 4894

RESULT 11  
US-09-071-710-16  
; Sequence 16, Application us/09071710  
; Patent No. 6130043





QY 2555 GCCTGTTGGTCTCTGTCATCACAGACACAGCAATTAATAATTTAACTATTT 2614  
Db 1381 GCCTGTTGGTCTCTGTCATCACAGACACAGCAATTAATAATTTAACTATTT 1440  
QY 2615 ATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTTCTGTTGGTGTCTAAATTTGG 2674  
Db 1441 ATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTTCTGTTGGTGTCTAAATTTGG 1500  
QY 2675 GTAGGTGGGGATCCCAACAAATCAGTCCCTGAGATAGTGTCTATTTGGGTGATCA 2734  
Db 1501 GTAGGTGGGGATCCCAACAAATCAGTCCCTGAGATAGTGTCTATTTGGGTGATCA 1560  
QY 2735 TTGCAGAAATCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2794  
Db 1561 TTGCAGAAATCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620  
QY 2795 AAATTTACTATCCCAATATTAATCCAAATGCTGTGTACCAAGTTAGGGTTGAA 2854  
Db 1621 AAATTTACTATCCCAATATTAATCCAAATGCTGTGTACCAAGTTAGGGTTGAA 1680  
QY 2855 GGAAGTAGAGGTGGGGTTCAGTCTCAAGGGTTCCTTAACCAAGTCTTCTCTG 2914  
Db 1681 GGAAGTAGAGGTGGGGTTCAGTCTCAAGGGTTCCTTAACCAAGTCTTCTCTG 1740  
QY 2915 GCCAGCTGTTTCCCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 2974  
Db 1741 GCCAGCTGTTTCCCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 1800  
QY 2975 AGSACTGCCCAAAATTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 3034  
Db 1801 AGSACTGCCCAAAATTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 1860  
QY 3035 GCTCCACACCTGTTTGGAGTCTGAGGACGAGACCAAAAGTGGGGTTTCCCAAG 3094  
Db 1861 GCTCCACACCTGTTTGGAGTCTGAGGACGAGACCAAAAGTGGGGTTTCCCAAG 1920  
QY 3095 CTTTCTGCTATCTCAGCCCCCAGATATCTGCTGGGAATCTCACAGAACTC 3154  
Db 1921 CTTTCTGCTATCTCAGCCCCCAGATATCTGCTGGGAATCTCACAGAACTC 1980  
QY 3155 AGGAGCACCCTGCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTCCG 3214  
Db 1981 AGGAGCACCCTGCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTCCG 2040  
QY 3215 TTGCAATATGCTGCTTATTTATGAGGGGTGAATTTTATATGTAAGTACGCA 3274  
Db 2041 TTGCAATATGCTGCTTATTTATGAGGGGTGAATTTTATATGTAAGTACGCA 2100  
QY 3275 ATCAGATATATGTTTATGTTGACAAATTAAGGCTTCTTATATGTTA 3326  
Db 2101 ATCAGATATATGTTTATGTTGACAAATTAAGGCTTCTTATATGTTA 2152

## RESULT 12

us-09-525-397-16

Sequence 16, Application US/09525397

Patent No. 6252047

GENERAL INFORMATION:

APPLICANT: BILLING-MEDEL, PATRICIA

APPLICANT: COHEN, MAURICE

APPLICANT: COLPITTS, TRACEY L.

APPLICANT: FRIEDMAN, PAULA N.

APPLICANT: GORDON, JULIAN

APPLICANT: GRANADOS, EDWARD N.

APPLICANT: HODGES, STEVEN C.

APPLICANT: KLASS, MICHAEL R.

APPLICANT: KRATOCHVIL, JON D.

APPLICANT: ROBERTS-RAPP, LISA

APPLICANT: RUSSELL, JOHN C.

APPLICANT: STROUPE, STEPHEN D.

TITLE OF INVENTION: REAGENTS AND METHODS USEFUL

FOR DETECTING DISEASES OF THE PROSTATE

## NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories

STREET: 100 Abbott Park Road

CITY: Abbott Park

STATE: IL

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/525,397

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/071,710

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Becker, Cheryl L.

REGISTRATION NUMBER: 35,441

REFERENCE/DOCKET NUMBER: 6083-US.P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 847/935-1729

TELEFAX: 847/938-2623

TELEX:

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 2152 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-525-397-16

Query Match

Best Local Similarity 62.7%; Score 2136.4; DB 4; Length 2152;

Matches 2149; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1177 GGGGCTGTACAGGGCGTGGCCAGAGCTGACCGGGGACCGAGCGCGGAGACACTATGA 1236  
Db 1 GGGGCTGTACAGGGCGTGGCCAGAGCTGACCGGGGACCGAGCGCGGAGACACTATGA 60  
QY 1237 TGAAGCGCTTCGGATGGGCGCTGCTCTGAGTGCGCCATCTCCCTGGTCTT 1296  
Db 61 TGAAGCGCTTCGGATGGGCGCTGCTCTGAGTGCGCCATCTCCCTGGTCTT 120  
QY 1297 CTCCTGTGTCATGACCGGCTGGTGACAGGATTCGGCACTCGAGCAGTCTATTTGGCCAG 1356  
Db 121 CTCCTGTGTCATGACCGGCTGGTGACAGGATTCGGCACTCGAGCAGTCTATTTGGCCAG 180  
QY 1357 TGTGCGAGCTTCCCTGTGGTGCGGGTGCCACATGCTGTCACACAGTGGCGGTGT 1416  
Db 181 TGTGCGAGCTTCCCTGTGGTGCGGGTGCCACATGCTGTCACACAGTGGCGGTGT 240  
QY 1417 GACAGCTTTCAGCGCCCTTCAGCGGTTTCCACCTTCTCAGCCTTCGAGATCTCCCTACAC 1476  
Db 241 GACAGCTTTCAGCGCCCTTCAGCGGTTTCCACCTTCTCAGCCTTCGAGATCTCCCTACAC 300  
QY 1477 ACTGGCCCTCCCTTCTTACACCGGGGAGAGAGTGTTCCTGCCCAATACCAGGGGACAC 1536  
Db 301 ACTGGCCCTCCCTTCTTACACCGGGGAGAGAGTGTTCCTGCCCAATACCAGGGGACAC 360  
QY 1537 TGGAGGTGCTAGCAGTGGAGGACAGCTGATGACAGCTTCTTCCAGCGCCCTTAAGCCCTGG 1596  
Db 361 TGGAGGTGCTAGCAGTGGAGGACAGCTGATGACAGCTTCTTCCAGCGCCCTTAAGCCCTGG 420  
QY 1597 AGCTCCCTTCCCTTAATGGACAGTGGGTGCTGGAGGAGTGGCTTCTCCCACTCCACACC 1656  
Db 421 AGCTCCCTTCCCTTAATGGACAGTGGGTGCTGGAGGAGTGGCTTCTCCCACTCCACACC 480  
QY 1657 CGCGCTCTCGGGGGCTCTGCTGCTGATGCTCTCCGTACGTGTGGTGGGTGAGCCAC 1716



FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: 08/850,713  
APPLICATION NUMBER: 02-MAY-1997  
FILING DATE: 02-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6083, US. P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2143 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-071-710-15

Query Match 62.08; Score 2114.8; DB 3; Length 2143;  
Best Local Similarity 99.88; Pred: No. 0;  
Matches 2139; Conservative 0; Mismatches 2; Indels 3; Gaps 2;  
QY 1185 ACCAGGCGCTGCCAGAGCTGAGCGGACCGGACCGGAGACACATATGATGAAGCGG 1244  
DB 1 ACCAGGCGCTGCCAGAGCTGAGCGGACCGGACCGGAGACACATATGATGAAGCGG 60  
QY 1245 TTCGGATGGCAGCGCTGGGGCTGTTCTGCAGTGCACATCTCCCTGGTCTCTCTCTG 1304  
DB 61 TTCGGATGGCAGCGCTGGGGCTGTTCTGCAGTGCACATCTCCCTGGTCTCTCTCTG 120  
QY 1305 TCATGACCGGCTGGTGCACCGATTTGGGACCTCGAGCAGTCTATTTGGCGAGTGGCAG 1364  
DB 121 TCATGACCGGCTGGTGCACCGATTTGGGACCTCGAGCAGTCTATTTGGCGAGTGGCAG 180  
QY 1365 CTTTCCCTGTGGCTGCGGCTGCCACATGCTGTCACAGTGTGGCGGTGGTGCACAGCTT 1424  
DB 181 CTTTCCCTGTGGCTGCGGCTGCCACATGCTGTCACAGTGTGGCGGTGGTGCACAGCTT 240  
QY 1425 CAGCGCGCTTCACCGGCTTCCCTTCAGCCCTGCAGATCTCCCTACACACTGGCCT 1484  
DB 241 CAGCGCGCTTCACCGGCTTCCCTTCAGCCCTGCAGATCTCCCTACACACTGGCCT 300  
QY 1485 CCCTCTACCGGAGAGAGAGTGTCTCTGCCCAATACCGAGGGACACTGGAGGTG 1544  
DB 301 CCCTCTACCGGAGAGAGAGTGTCTCTGCCCAATACCGAGGGACACTGGAGGTG 360  
QY 1545 CTAGCAGTGAAGGACAGCTGATGACAGCTTCTGTCAGGCGCTTAAGCCCTGGAGCTCCCT 1604  
DB 361 CTAGCAGTGAAGGACAGCTGATGACAGCTTCTGTCAGGCGCTTAAGCCCTGGAGCTCCCT 420  
QY 1605 TCCCTAATGACAGCTGGGTGCTGGAGCAGTGGCTGTCCACCTTCCACCGCGCTCT 1664  
DB 421 TCCCTAATGACAGCTGGGTGCTGGAGCAGTGGCTGTCCACCTTCCACCGCGCTCT 480  
QY 1665 CGGGGGCTCTGCTGTGATGTCTCCGTAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1724  
DB 481 CGGGGGCTCTGCTGTGATGTCTCCGTAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 540  
QY 1725 GGTGGTTCGGGGCGGGGATCTGCTGGACCTTCCGACCTTCCGATCTGCTGCTCTCTG 1784  
DB 541 GGTGGTTCGGGGCGGGGATCTGCTGGACCTTCCGACCTTCCGATCTGCTGCTCTCTG 600  
QY 1785 TGTCCAGGTGGCGGCTTCCCTGTTATGGGCTCCATTGCGGCTCAGGCTCAGGCTGTGCTCA 1844  
DB 601 TGTCCAGGTGGCGGCTTCCCTGTTATGGGCTCCATTGCGGCTCAGGCTCAGGCTGTGCTCA 660  
QY 1845 CTGCTATATGCTGTGCGCGAGCGCTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTG 1904  
DB 661 CTGCTATATGCTGTGCGCGAGCGCTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTG 720

QY 1905 TAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACCTTCCAGCACATGG 1964  
DB 721 TAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACCTTCCAGCACATGG 780  
QY 1965 GGTGAGGGCTGCTCCTCACTGGTCCAGCTCCCGCTCCCTCTGTTAGCCCCATGGGGCTGC 2024  
DB 781 GGTGAGGGCTGCTCCTCACTGGTCCAGCTCCCGCTCCCTCTGTTAGCCCCATGGGGCTGC 840  
QY 2025 CGGGCTGGCGCCAGTCTGTTGCTGCCAAAGTAATGTGCTCTCTCTCTCTCTCTCTCT 2084  
DB 841 CGGGCTGGCGCCAGTCTGTTGCTGCCAAAGTAATGTGCTCTCTCTCTCTCTCTCTCTCT 900  
QY 2085 GCTGCTGAGTGGCTGAGTGCACAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGG 2144  
DB 901 GCTGCTGAGTGGCTGAGTGCACAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGG 960  
QY 2145 TCTTAGGGCTGCTGAGTGGAGGCTTCCAAAGGGGGTTCAGTGTGAGCTTATACAGGG 2204  
DB 961 TCTTAGGGCTGCTGAGTGGAGGCTTCCAAAGGGGGTTCAGTGTGAGCTTATACAGGG 1020  
QY 2205 AGGCCAGAGGGCTCCATGCACTGGAATGCCGGGACTCTGCAGGTGGATTACCCAGGCTC 2264  
DB 1021 AGGCCAGAGGGCTCCATGCACTGGAATGCGGGACTCTGCAGGTGGATTACCCAGGCTC 1079  
QY 2265 AGGGTTAACAGCTAGCTCCTAGTTGAGACACACACTAGAGAGGGTTCCTGGAGCTGAA 2324  
DB 1080 AGGGTTAACAGCTAGCTCCTAGTTGAGACACACACTAGAGAGGGTTCCTGGAGCTGAA 1139  
QY 2325 TAACTCAGTCACTGGTTCCTCTTAAGCCCCCTTAACCTGCAGCTTCGTTTAAATGT 2384  
DB 1140 TAACTCAGTCACTGGTTCCTCTTAAGCCCCCTTAACCTGCAGCTTCGTTTAAATGT 1199  
QY 2385 AGCTCTGATGGGAGTTCCTAGGATGAACACTCTCCATGGGATTTGAACATATGAA 1259  
DB 1200 AGCTCTGATGGGAGTTCCTAGGATGAACACTCTCCATGGGATTTGAACATATGAA 1259  
QY 2443 CTTATTTGAGGGGAGAGTCTCTGAGGGGCAACACACAGAGTCCCTCAGGCCA 2502  
DB 1260 GTTATTTGAGGGGAGAGTCTCTGAGGGGCAACACACAGAGTCCCTCAGGCCA 1319  
QY 2503 CAGCAGTGTCTTTTGTGATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTG 2562  
DB 1320 CAGCAGTGTCTTTTGTGATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTG 1379  
QY 2563 GTCTTCTGTTCCCATCAGAGACAGAGGCTTTAAATATTTAACTTATTTTAACTA 2622  
DB 1380 GTCTTCTGTTCCCATCAGAGACAGAGGCTTTAAATATTTAACTTATTTTAACTA 1439  
QY 2623 AAGTAGAGGGAATCCATTGCTAGCTTTTCTGTTGGTGTCTAATATTTGGGTAGGGTG 2682  
DB 1440 AAGTAGAGGGAATCCATTGCTAGCTTTTCTGTTGGTGTCTAATATTTGGGTAGGGTG 1499  
QY 2683 GGGGATCCCAACATCAGGTCCCTGAGATAGCTTGGCTGATCATTTCCAGG 2742  
DB 1500 GGGGATCCCAACATCAGGTCCCTGAGATAGCTTGGCTGATCATTTCCAGG 1559  
QY 2743 ATCTTCTCTCTGGGCTGCTGCCCCCAAAATGCTTAACCCAGGCTTGGAAATCTTA 2802  
DB 1560 ATCTTCTCTCTGGGCTGCTGCCCCCAAAATGCTTAACCCAGGCTTGGAAATCTTA 1619  
QY 2803 CTCATCCCAAAATGATAATTTCCAAATGCTTTACCCAAAGGTAGGGTGTGAAGGAGTA 2862  
DB 1620 CTCATCCCAAAATGATAATTTCCAAATGCTTTACCCAAAGGTAGGGTGTGAAGGAGTA 1679  
QY 2863 GAGGTGGGGCTTCAGGTCTCAACGGCTTCCCTTAACCCAGGCTTCTCTCTTCTGGCCAGCC 2922  
DB 1680 GAGGTGGGGCTTCAGGTCTCAACGGCTTCCCTTAACCCAGGCTTCTCTCTTCTGGCCAGCC 1739  
QY 2923 TGGTTCCTCCCTCTTCCACTCCCTCTACTCTCTAGGCTGGCTGATGAGGACTG 2982  
DB 1740 TGGTTCCTCCCTCTTCCACTCCCTCTACTCTCTAGGCTGGCTGATGAGGACTG 1799  
QY 2983 CCAAAATTTCCCTACCCCAAACTTTCCCTTACCCCAAACTTTCCCAACAGCTCCACA 3042

[illegible]

## RESULT 14

US-09-525-397-15  
Sequence 15, Application US/09525397  
Patent No. 6252047  
GENERAL INFORMATION:  
APPLICANT: BILLING-MEDEL, PATRICIA  
APPLICANT: COHEN, MAURICE  
APPLICANT: COLPITTS, TRACEY L.  
APPLICANT: FRIEDMAN, PAULA N.  
APPLICANT: GORDON, JULIAN  
APPLICANT: GRANADOS, EDWARD N.  
APPLICANT: HODGES, STEVEN C.  
APPLICANT: KLASS, MICHAEL R.  
APPLICANT: KRATOCHVIL, JON D.  
APPLICANT: ROBERTS-RAPP, LISA  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STROUPE, STEPHEN D.  
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
FOR DETECTING DISEASES OF THE PROSTATE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/525,397  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/071,710  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6083.US.P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:









Db 599 CGTAAATTCAGACATTTGGGGTGGAGGCTTCCCTACTGGGT-CCAACTCCC 657  
Qy 1999 CGCTCTCTAGCCCATGGGCTGCGGGCTGGCGGCGGAGTTCTGTGTGCTGCCAAAGT 2058  
Db 658 CGCTCTCTTAACCCCATGGGCTGCGGGCTTGGCGGCGCAATTTCTGTGTGCTGCCAAAT 717  
Qy 2059 AATGTGGCTCTGCTGCGCACCTCTGTCT-GCTGAGGTGCTGAGCTGCACAGCTGGGGC 2117  
Db 718 NATGTGGCTCTGCTGCGCACCTGTGTCTGTGCTGAGTGCNACNCTNGGGGG 777  
Qy 2118 TGGGGCTGCC 2128  
Db 778 TNGGGNGTCC 788

## RESULT 18

US-09-605-785-10  
; Sequence 10, Application US/09605785  
; Patent No. 6321716

; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqi

; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.  
; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel  
; APPLICANT: Wang, Aljun

; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.427C16

; CURRENT APPLICATION NUMBER: US/09/605,785  
; CURRENT FILING DATE: 2000-06-27

; NUMBER OF SEQ ID NOS: 835  
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 10  
; LENGTH: 789

; TYPE: DNA  
; ORGANISM: Homo sapien

; FEATURE:  
; NAME/KEY: misc\_feature

; LOCATION: (1)...(789)  
; OTHER INFORMATION: n = A,T,C or G

US-09-605-785-10

Query Match 19.78; Score 673.4; DB 4; Length 789;  
Best Local Similarity 94.28; Pred. No. 2.9e-124;  
Matches 745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;

Qy 1341 CAGTCTATTGGCCAGTGGCAGCTTCCCTGTGGCTGCGGTCGCACATGCTGCC 1400  
Db 1 CAGTCTATNNGCCAGTGGCAGCTTCCCTGTGGCTGCGGTCGCACATGCTGCC 60  
Qy 1401 ACAGTGTGGCCGTGGTGCACAGTTCACGCCCTTACCGGGTTCACCTTCTAGCCCTGC 1460  
Db 61 ACAGTGTGGCCGTGGTGCACAGTTCACGCCCTTACCGGGTTCACCTTCTAGCCCTGC 120  
Qy 1461 AGATCTGCCCTACACACTGGCTTCCCTTACACCGGGAGAGCAGTCTTCTGCCCA 1520  
Db 121 AGATCTGCCCTACACACTGGCTTCCCTTACACCGGGAGAGCAGTCTTCTGCCCA 180  
Qy 1521 AATACCGGGGACACTGGAGTGTCTAGCAGTGGAGCAGCTGATGACAGCTTCTCTGC 1580

Db 181 AATACCAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCGTGATGACACAGCTTCTCTGC 240  
Qy 1581 CAGGCCCTAAGCCCTGGAGCTCCCTTCCCTAATGACACAGTGGGTGCTGGAGGAGTGGCC 1640  
Db 241 CAGGCCCTAAGCCCTGGAGCTCCCTTCCCTAATGACACAGTGGGTGCTGGAGGAGTGGCC 300  
Qy 1641 TGTCTCCACCTCCACCCGCGCTCTGCGGGGCTCTGCTGTGATGTCTCCGTACGTGTGG 1700  
Db 301 TGTCTCCACCTCCACCCGCGCTCTGCGGGGCTCTGCTGTGATGTCTCCGTACGTGTGG 360  
Qy 1701 TGTGGGTGAGCCACCGAGGAGTGGTTCGGGCGGGGCGATCTGCTTGACCTCG 1760  
Db 361 TGTGGGTGAGCCACCGAGGAGTGGTTCGGGCGGGGCGATCTGCTTGACCTCG 420  
Qy 1761 CCATCTCTGGATAGTGTCTGCTGCTCCAGGTGGCCCATCCCTGTTTATGGGTCCA 1820  
Db 421 CCATCTCTGGATAGTGTCTGCTGCTCCAGGTGGCCCATCCCTGTTTATGGGTCCA 479  
Qy 1821 TTGTCCAGCTCAGCCAGTGTCTGCTGCTATATGTTGCTGCGGCGAGGCTGGGTCTGG 1880  
Db 480 TTGTCCAGCTCAGCCAGTGTCTGCTGCTATATGTTGCTGCGGCGAGGCTGGGTCTGG 539  
Qy 1881 TCGCCATTCTTCTTGTACACAGTGTATTTGACAAGAGGAGTGTGGCCAAATACCTCAG 1940  
Db 540 TC-CCATTCTTGTGTACACAGTGTATTTGACAAGAGGAGTGTGGCCAAATACCTCAG 598  
Qy 1941 CGTAGAAACTTCCAGCA--CATTGGGTGGAGGCGCTGCTCAGTGGTCCAGCTCCC 1998  
Db 599 CGTTAAAAAATTCAGCAACATTTGGGGGTGGAAGGCTTGCCTCAGTGGT-CCAACTCCC 657  
Qy 1999 CGTCTCTTTAGCCCCCATGGGCTGCGGGCTGCGGCGAGTTCCTGTGCTGCCAAAGT 2058  
Db 658 CGTCTCTTTAAACCCCATGGGCTGCGGGCTGCGGCGCAATTTCTGTGCTGCCAAANT 717  
Qy 2059 AATGTGGCTCTGCTGCCACCTGTGCT-GCTGAGGTGCTGAGTGTGACAGCTGGGGC 2117  
Db 718 NATGTGGCTCTGCTGCTGCCACCTGTGCTGCTGAGTGCNACNCTNGGGGG 777  
Qy 2118 TGGGGCTGCC 2128  
Db 778 TNGGGNGTCC 788

## RESULT 19

US-09-439-313-10  
; Sequence 10, Application US/09439313  
; Patent No. 6329505

; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang Yuqi

; APPLICANT: Reed, Steven G.  
; APPLICANT: Kalos, Michael

; APPLICANT: Fanger, Gary  
; APPLICANT: Retter, Mark

; APPLICANT: Solk, John  
; APPLICANT: Day, Craig

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
; FILE REFERENCE: 210121.427C9

; CURRENT APPLICATION NUMBER: US/09/439,313  
; CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 575  
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 10  
; LENGTH: 789

; TYPE: DNA  
; ORGANISM: Homo sapien

; FEATURE:  
; NAME/KEY: misc\_feature

; LOCATION: (1)...(789)

; OTHER INFORMATION: n = A, T, C or G  
US-09-439-313-10

Query Match	19.7%	Score 673.4	DB 4	Length 789
Best Local Similarity	94.2%	Pred. No. 2.9e-124		
Matches 745; Conservative	0	Mismatches 40	Indels	Gaps 5
QY	1341	CAGTCTATTGGCCAGGTGGCAGCTTCCCTGTGGCTGCGGTGCCACATGCTGTCCC	1400	
DB	1	CAGTCTATNTGGCCAGGTGGCAGCTTCCCTGTGGCTGCGGTGCCACATGCTGTCCC	60	
QY	1401	ACAGTGTGGCGGTGGTGACAGCTTCACGCCCCCTCACCGGTTTCACCTTCTCAGCCCTGC	1460	
DB	61	ACAGTGTGGCGGTGGTGACAGCTTCACGCCCCCTCACCGGTTTCACCTTCTCAGCCCTGC	120	
QY	1461	AGATCCCTGCCCTACACACTGGCCTCCCTCTACACCGGGAGAGCAGGTGTTCCCTGCCCA	1520	
DB	121	AGATCCCTGCCCTACACACTGGCCTCCCTCTACACCGGGAGAGCAGGTGTTCCCTGCCCA	180	
QY	1521	AATACCAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCACTTCTCTGC	1580	
DB	181	AATACCAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCACTTCTCTGC	240	
QY	1581	CAGCCCTAAGCTTGGAGCTCCCTTCCTTAATGGACACAGTGGGTGTGGAGGCAGTGGCC	1640	
DB	241	CAGCCCTAAGCTTGGAGCTCCCTTCCTTAATGGACACAGTGGGTGTGGAGGCAGTGGCC	300	
QY	1641	TGCTCCCACTCCACCGCGCTCTGGGGGCGCTTGCTCTGATGTCCTCGTACGTGTGG	1700	
DB	301	TGCTCCCACTCCACCGCGCTCTGGGGGCGCTTGCTCTGATGTCCTCGTACGTGTGG	360	
QY	1701	TGTTGGGTGAGCCACCGAGGCCAGGTGGTCCGGGCGGGGCATCTGCTGGACCTCG	1760	
DB	361	TGTTGGGTGAGCCACCGAGGCCAGGTGGTCCGGGCGGGGCATCTGCTGGACCTCG	420	
QY	1761	CCATCTCGGATAGTGGCTTCTGCTGCCAGGTGGCCCATCCCTGTTTATGGGTCCA	1820	
DB	421	CCATCTCGGATAGT - CTTCTGCTGTGCCANGTGGCCCATCCCTGTTTATGGGTCCA	479	
QY	1821	TTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTCTGCCGAGGCGCTGGGTCTGG	1880	
DB	480	TTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTCTGCCGAGGCGCTGGGTCTGG	539	
QY	1881	TCGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATCTCAG	1940	
DB	540	TC - CCAATTTACTTTGCTACAGGTAATATTTGACAAGAACGANTTTGGCCAAATCTCAG	598	
QY	1941	CGTAGAAAACCTCCAGCA - - CATTTGGGTGGAGGGCTTGCTACATGGGTCCAGCTCCC	1998	
DB	599	CGTTAAAAAATTCAGCAACATTTGGGGTGGAGGCGCTGCTCAGTGGGT - CCAACTCCC	657	
QY	1999	CGCTCTGTTTAGCCCATGGGGCTGCCGGGCTGGCCGCAATTTCTGTTCTGCTGCCAAAGT	2058	
DB	658	CGCTCTGTTTAAACCATGGGGCTGCCGGCTTGCCGCGCAATTTCTGTTGCTGCCAAANT	717	
QY	2059	AATGTGGCTCTGCTGCCACCTGTGCT - GCTGAGGTGGGTAGCTGCACAGCTGGGGC	2117	
DB	718	NATGTGGCTCTGCTGCCACCTGTTTGGTGGCTGAAAGTGCNTACNCGCNACNTGGGGG	777	
QY	2118	TGGGGCGTCCC	2128	
DB	778	TNGGGNGTCC	788	

RESULT 20

US-09-352-616A-10  
 : Sequence 10, Application US/09352616A  
 : Patent No. 6395278  
 : GENERAL INFORMATION:  
 : APPLICANT: Dillon, Davin C.  
 : APPLICANT: Harlocker, Susan Louise  
 : APPLICANT: Uiang, Yuqi  
 : APPLICANT: Xu, Jiangchun





[illegible][illegible]











APPLICANT: ROBERTS-RAPP, LISA  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STROUPE, STEPHEN D.  
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
FOR DETECTING DISEASES OF THE PROSTATE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,710  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/850,713  
FILING DATE: 02-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6083.US.P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:

Sequence 10, Application US/09525397  
Patent No. 6252047  
GENERAL INFORMATION:  
APPLICANT: BILLING-MEDEL, PATRICIA  
APPLICANT: COHEN, MAURICE  
APPLICANT: COLPITTS, TRACEY L.  
APPLICANT: FRIEDMAN, PAULA N.  
APPLICANT: GORDON, JULIAN  
APPLICANT: GRANADOS, EDWARD N.  
APPLICANT: HODGES, STEVEN C.  
APPLICANT: KLAS, MICHAEL R.  
APPLICANT: KRATOCHVIL, JON D.  
APPLICANT: ROBERTS-RAPP, LISA  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STROUPE, STEPHEN D.  
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
FOR DETECTING DISEASES OF THE PROSTATE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/525,397  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/071,710  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6083.US.P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 288 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: base\_polymorphism  
LOCATION: 147  
OTHER INFORMATION: /note= " N' represents an A or G or  
OTHER INFORMATION: T or C polymorphism at this position"  
US-09-525-397-10

Query Match 8.4%; Score 287; DB 4; Length 288;  
Best Local Similarity 99.7%; Pred. No. 3.8e-48;  
Matches 287; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2532 CTCCTACCTTTTATCAGGATGGCGCTGTTGGTCTCTCTGTTGCCATCACAGACACAG 2591  
Db 1 CTCCTACCTTTTATCAGGATGGCGCTGTTGGTCTCTCTGTTGCCATCACAGACACAG 60

QY 2592 GCATTAAATATTAACTATTATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTT 2651  
Db 61 GCATTAAATATTAACTATTATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTT 120

QY 2652 CTGCTGTTGGTCTTAATATTATTGGTAGGTGGGGATCCCAACAAATCAGTCCCTGAG 2711  
Db 121 CTGCTGTTGGTCTTAATATTATTGGTAGGTGGGGATCCCAACAAATCAGTCCCTGAG 180

RESULT 33  
US-09-525-397-10

Query Match 8.4%; Score 287; DB 3; Length 288;  
Best Local Similarity 99.7%; Pred. No. 3.8e-48;  
Matches 287; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2532 CTCCTACCTTTTATCAGGATGGCGCTGTTGGTCTCTCTGTTGCCATCACAGACACAG 2591  
Db 1 CTCCTACCTTTTATCAGGATGGCGCTGTTGGTCTCTCTGTTGCCATCACAGACACAG 60

QY 2592 GCATTAAATATTAACTATTATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTT 2651  
Db 61 GCATTAAATATTAACTATTATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTT 120

QY 2652 CTGCTGTTGGTCTTAATATTATTGGTAGGTGGGGATCCCAACAAATCAGTCCCTGAG 2711  
Db 121 CTGCTGTTGGTCTTAATATTATTGGTAGGTGGGGATCCCAACAAATCAGTCCCTGAG 180

QY 2712 ATAGCTGGTCAATGGGCTGATCATGTCAGAACTTTCTCTCTGGGGTCTGGCCCCCA 2771  
Db 181 ATAGCTGGTCAATGGGCTGATCATGTCAGAACTTTCTCTCTGGGGTCTGGCCCCCA 240

QY 2772 AAATGCCTAACCCAGGACCTTGGAAATTTACTCATCCCAAAATGATAA 2819  
Db 241 AAATGCCTAACCCAGGACCTTGGAAATTTACTCATCCCAAAATGATAA 288

RESULT 33  
US-09-525-397-10

QY 2712 ATAGCTGGTCATTTGGGCTGATTCATGGCCAGAAATCTTCTCTCCCTGGGCTCGGCCCCCA 2771  
DB 181 ATAGCTGGTCATTTGGGCTGATTCATGGCCAGAAATCTTCTCTCCCTGGGCTCGGCCCCCA 240  
QY 2772 AAATGCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCAAAATGATAA 2819  
DB 241 AAATGCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCAAAATGATAA 288

## RESULT 34

US-09-071-710-11  
; Sequence 11, Application US/09071710  
; Patent No. 6130043  
; GENERAL INFORMATION:  
; APPLICANT: BILLING-MEDEL, PATRICIA  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GORDON, JULIAN  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: HODGES, STEVEN C.  
; APPLICANT: KLASS, MICHAEL R.  
; APPLICANT: KRATOCHVIL, JON D.  
; APPLICANT: ROBERTS-RAPP, LISA  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STROUPE, STEPHEN D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; FOR DETECTING DISEASES OF THE PROSTATE  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,710  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/850,713  
; FILING DATE: 02-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6083.US.P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 272 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; NAME/KEY: base\_polymorphism  
; LOCATION: 216  
; OTHER INFORMATION: /note= " N' represents an A or G or  
; OTHER INFORMATION: T or C polymorphism at this position"  
US-09-071-710-11

Query Match 7.9%; Score 270; DB 3; Length 272;  
Best Local Similarity 99.6%; Pred. No. 8.4e-45;  
Matches 270; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2795 AAATTTCTACTCATCCCAAAATGATAATTCCAAATGCTGTTTACCCAAAGTTAGGGTGTGAA 2854  
DB 1 AAATTTCTACTCATCCCAAAATGATAATTCCAAATGCTGTTTACCCAAAGTTAGGGTGTGAA 60  
QY 2855 GGAAGGTAGAGGTGGGGCTTCAGGTCTCAACGGCTTCCCTAACCAACCCCTCTTCTCTTG 2914  
DB 61 GGAAGGTAGAGGTGGGGCTTCAGGTCTCAACGGCTTCCCTAACCAACCCCTCTTCTCTTG 120  
QY 2915 GCCCAGCCTGTTCCCGCCCACTTCCCACTCCCTCTACTCTCTAGGACTGGGCTCATGA 2974  
DB 121 GCCCAGCCTGTTCCCGCCCACTTCCCACTCCCTCTACTCTCTAGGACTGGGCTCATGA 180  
QY 2975 AGGCACTGCCCAAAATTTCCCTTACCCCACTTCCCTTACCCCACTTCCCTTACCCCACT 3034  
DB 181 AGGCACTGCCCAAAATTTCCCTTACCCCACTTCCCTTACCCCACTTCCCTTACCCCACT 240  
QY 3035 GCTCCACAACCCCTGTTGGAGCTACTGCAGG 3065  
DB 241 GCTCCACAACCCCTGTTGGAGCTACTGCAGG 271

## RESULT 35

US-09-325-397-11  
; Sequence 11, Application US/09525397  
; Patent No. 6252047  
; GENERAL INFORMATION:  
; APPLICANT: BILLING-MEDEL, PATRICIA  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GORDON, JULIAN  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: HODGES, STEVEN C.  
; APPLICANT: KLASS, MICHAEL R.  
; APPLICANT: KRATOCHVIL, JON D.  
; APPLICANT: ROBERTS-RAPP, LISA  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STROUPE, STEPHEN D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; FOR DETECTING DISEASES OF THE PROSTATE  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/525,397  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/071,710  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6083.US.P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 272 base pairs  
; TYPE: nucleic acid





```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/525,397
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,710
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083.US.P1
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-525-397-9

Query Match          7.8%; Score 265; DB 4; Length 265;
Best Local Similarity 100.0%; Pred. No. 8.1e-44;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2453 GGGGAAGAGTCTCTGAGGGGCAACACACAGAACACAGGTCCCTCAGCCACAGCACTGTC 2512
DB 1 GGGGAAGAGTCTCTGAGGGGCAACACACAGAACACAGGTCCCTCAGCCACAGCACTGTC 60

QY 2513 TTTTGTCTGATCCACCCCTCTTACCTTTATCAGGATGTGGCTGTGTCCTCTGT 2572
DB 61 TTTTGTCTGATCCACCCCTCTTACCTTTATCAGGATGTGGCTGTGTCCTCTGT 120

QY 2573 TGGCATCAGACAGACAGGCAATTAATATTTAACTTATTTATCAAGTAGAAGG 2632
DB 121 TGGCATCAGACAGACAGGCAATTTAAATATTTAACTTATTTATCAAGTAGAAGG 180

QY 2633 GAATCATTGCTAGCTTTTCTGTGTGTTCTAATATTTGGTAGGGTGGGGATCCCC 2692
DB 181 GAATCATTGCTAGCTTTTCTGTGTGTTCTAATATTTGGTAGGGTGGGGATCCCC 240

QY 2693 ACAATCAGGTCCCTGAGATAGCT 2717
DB 241 ACAATCAGGTCCCTGAGATAGCT 265

RESULT 38
US-09-071-710-1
; Sequence 1, Application US/09071710
; Patent No. 6130043
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; FOR DETECTING DISEASES OF THE PROSTATE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESS: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
```

```
;
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,710
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/850,713
; FILING DATE: 02-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083.US.P1
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-071-710-1

Query Match          7.5%; Score 256.4; DB 3; Length 258;
Best Local Similarity 99.6%; Pred. No. 4e-42;
Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1177 GGGGCTCTACAGGGCGTGCAGAGCTGAGCCGGGACCGGCGGACCGGAGACACTATGA 1236
DB 1 GGGGCTCTACAGGGCGTGCAGAGCTGAGCCGGGACCGGCGGACCGGAGACACTATGA 60

QY 1237 TGAAGGGGTTGCGATGGGAGCCCTGGGGCTGTTCTTCAGTGGCCATCTCCCTGGTCTT 1296
DB 61 TGAAGGGGTTGCGATGGGAGCCCTGGGGCTGTTCTTCAGTGGCCATCTCCCTGGTCTT 120

QY 1297 CTCTCTGTATGACGCGGCTGGTGGCAGGATTCGGGACCTCGAGCAGTCTATTGGCCAG 1356
DB 121 CTCTCTGTATGACGCGGCTGGTGGCAGGATTCGGGACCTCGAGCAGTCTATTGGCCAG 180

QY 1357 TGTGGCAGCTTTCCTGTGCTGCCGTGCCGTGCACATGCTGTCCACAGTGTGGCCGTGGT 1416
DB 181 TGTGGCAGCTTTCCTGTGCTGCCGTGCCGTGCACATGCTGTCCACAGTGTGGCCGTGGT 240

QY 1417 GACAGCTTCAGCCGCCCT 1434
DB 241 GACAGCTTCAGCCGCCCT 258

RESULT 39
US-09-525-397-1
; Sequence 1, Application US/09525397
; Patent No. 6252047
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
```







RESULT 45  
US-09-071-710-5  
; Sequence 5, Application US/09071710  
; Patent No. 6130043  
; GENERAL INFORMATION:  
; APPLICANT: BILLING-MEDEL, PATRICIA  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GORDON, JULIAN  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: HODGES, STEVEN C.  
; APPLICANT: KLASS, MICHAEL R.  
; APPLICANT: KRATOCHVIL, JON D.  
; APPLICANT: ROBERTS-RAPP, LISA  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STROUPE, STEPHEN D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; FOR DETECTING DISEASES OF THE PROSTATE  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071.710  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/850.713  
; FILING DATE: 02-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6083.US.P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 231 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-09-071-710-5

Query Match 6.5%; Score 220; DB 3; Length 231;  
Best Local Similarity 99.6%; Pred. No. 5.9e-35;  
Matches 231; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1767 TGGATAGTGCCTCTCTGCTGCCAGGTGGCCCATCCCTGTTATGGGCTCCATTGTCC 1826  
Db 1 TGGATAGTGCCTCTCTGCTGCCAGGTGGCCCATCCCTGTTATGGGCTCCATTGTCC 60  
QY 1827 AGCTCAGCCACTGTCTACTGCCCTATATGTTCTGCCGAGGCGCTGGTCTGTGGCCA 1886  
Db 61 AGCTCAGCCAGTGTCTACTGCCCTATATGTTCTGCCGAGG -CTGGTCTGTGGCCA 119  
QY 1887 TTTACTTTGCTACACAGGTAGTATTGACAGAGGAGCTTGGCCAAATACACGGGTAGA 1946  
Db 120 TTTACTTTGCTACACAGGTAGTATTGACAGAGGAGCTTGGCCAAATACACGGGTAGA 179  
QY 1947 AAACCTCCAGCACATTGGGGTGGAGGGGCTGCCTCAGCTGGGTCCCGCTCCC 1998



\_\_\_\_\_

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	496	7.7	530	1	MATP_HUMAN	Q9umx9 homo sapien
2	489.5	7.6	530	1	MATP_MOUSE	P58355 mus musculu
C 3	398.5	6.4	660	1	YHL1_EBV	P03181 Epstein-Bar
C 4	373	6.0	1453	1	CAL1_MOUSE	P11087 mus musculu
C 5	358.5	5.8	1464	1	CAL3_MOUSE	P08421 mus musculu
C 6	352.5	5.7	1262	1	CAL1_CHICK	P12105 gallus gall
C 7	351.5	5.6	1466	1	CAL3_HUMAN	P20461 homo sapien
C 8	350	5.6	1460	1	CAL1_CANFA	Q9x3j7 canis fami
C 9	348	5.6	1690	1	C4A4_HUMAN	P53420 homo sapien
C 10	344.5	5.5	1049	1	CAL1_BOVIN	P04258 bos taurus
C 11	344	5.5	1453	1	CAL1_CHICK	P02457 gallus gall
C 12	343	5.5	1464	1	CAL1_CHICK	P02452 homo sapien
C 13	337.5	5.4	1838	1	CAL1_HUMAN	P20908 homo sapien
C 14	330	5.3	1603	1	CALF_HUMAN	Q07092 homo sapien
C 15	327	5.1	525	1	STP_SP10L	Q03411 spinacia ol
C 16	326	5.2	1459	1	CAL2_MOUSE	P28481 mus musculu
C 17	324.5	5.2	1355	1	C421_RANCA	O42350 rana catesb
C 18	324.5	5.2	1418	1	CAL2_HUMAN	P02458 homo sapien





```

QY 512 CACTGGCGTGGACGCTATAGCCCGCCGCGCCCTTCACTGGGCATCTCTTGGGCATC 571
Db 94 HisCysArgAlaArgTrpGlyArgArgProTyrIleLeuThrLeuAlaIleMetMet 113
QY 572 CTGCTAGCCTCTTCTCATCCCAAGGCGCGCTGCTAGCAGGGCTGCTGCCCGGAT 631
Db 114 LeuLeuGlyMetAlaLeuTyrLeuAsnGlyAspAlaValValSerAlaLeuValAlaAsn 133
QY 632 CCCAGGCCC-----CTGGAGCTGGCCTCATCTCTGGCGTGGGGTGGTGGAC 682
Db 134 ProArgGlnLysLeuIleTyrAlaIleSerIleThrMetValGlyValValLeuPheAsp 153
QY 683 TTTCTGGGCGAGTGTCTTCACTCCACTGGAGCCCTGCTCTCTCACTCTCTCCGGGAC 742
Db 154 PheSerAlaAspPheIleAspGlyProIleLysAlaTyrLeuPheAspValCysSerHis 173
QY 743 CCGGACCACTGTCGCCAGGCTACTCTGTATGCTTCTATGATCATGCTTGGGGCTGC 802
Db 174 GlnAspLys---GluLysGlyLeuHisTyrHisAlaLeuPheThrGlyPheGlyAla 192
QY 803 CTGGCTTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 862
Db 193 LeuGlyTyrIleLeuGlyAlaIleAspTrpValHisLeuAspLeuGlyArgLeuLeuGly 212
QY 863 ACCGAGGAGGAGTGTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 922
Db 213 ThrGluPheGlnValMetPhePheSerAlaLeuValLeuIleLeuGlyCysPheIleThr 232
QY 923 ACAGTGTGCTGCTGAGGAGGAGGCTG-----GGCCCCACCGAGCA----- 967
Db 233 HisLeuCysSerIleProGluAlaProLeuArgAspAlaThrAspProProSerGln 252
QY 968 -----GCAGAGGCTGCTGCGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1015
Db 253 GlnAspProGlnGlySerSerLeuSerAlaSerGlyMethHisGluTyr----- 268
QY 1016 CGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1048
Db 269 GlySerIleGluLysValLysAsnGlyGlyAlaAspThrGluGlnProValGlnGluTrp 288
QY 1049 -----CTGCTTCCCGGCTGCAACG 1069
Db 289 LysAsnLysLysProSerGlyGlnSerGlnArgThrMetSerMetLysSerLeuLeuArg 308
QY 1070 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1129
Db 309 AlaLeuValAsnMetProSerHisTyrArgCysLeuValSerHisLeuIleGlyTyr 328
QY 1130 ATGGCACTATGACCTTACGCTGCTTTACACGATTTCTGCGGCGAGGCTGTACCAG 1189
Db 329 ThrAlaPheLeuSerAsnMetLeuPhePheThrAspPheMetGlyGlnIleValTyrHis 348
QY 1190 GCGCTGCGCAGAGCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1249
Db 349 GlyAspProTyrGlyAlaHisAsnSerThrGluPheLeuIleTyrGluArgGlyValGlu 368
QY 1250 ATGGCGACCTGCGGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309
Db 369 ValGlyCysTrpGlyLeuCysIleAsnSerValPheSerSerValTyrSerTyrPheGln 388
QY 1310 GACCGGCTGTCGACGATTCGCGACCTGCGAGCTGCTATTGGCCAGTGTGGCGCTTTC 1369
Db 389 LysAlaMetValSerTyrIleGlyLeuLysGlySerPheMetGlyTyrLeuLeuPhe 408
QY 1370 CCGTGTGCTGCGGCTGCGCAGATGCTGCTGCGCAGAGTGTGCGGCTGCTGCTGCTGCT 1429
Db 409 GlyLeuGlyThrGlyPheIleGlyLeuPheProAsnValTyrSerThrLeuValLeuCys 428
QY 1430 GCCCTACCGGGTTCACCTTCTAGCCCTGAGAGTCTGCGCTGCTGCTGCTGCTGCTGCT 1489
Db 429 SerMetPheGlyValMetSerSerThrLeuTyrThrValProPheAsnLeuIleAlaGlu 448
QY 1490 TACCACCGGGAGAGCAGGTGTTCTGCCCAATACCGAGGAGGAGGAGGAGGAGGAGGAG 1549

```

```

Db 449 TyrHisArgGluGluGlu-----LysGluLysGlyGlnGluAla----- 461
QY 1550 AGTAGGACAGCCTGATCACCAGCTTCTGCGCAGGCGCTTAAGCCTGGAGCTCCCTCCCT 1609
Db 462 -----ProGlyGlyProAspAsn 467
QY 1610 AATGGACAGTGGTGGTGGAGGAGTGGCTGTGCTGCCACCTCCACCGCGCTCTGGGG 1659
Db 468 GlnGlyArg-----GlyLysGlyVal-----AspCysAla 477
QY 1670 GCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1729
Db 478 AlaLeuThrCysMetValGlnLeu-----AlaGlnIle 488
QY 1730 GTTCGGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
Db 489 LeuValGlyGly----- 493
QY 1790 CAGTGGCGCCCATCTGTTTATGGGCTCCATTGTCAGCTCAGCAGCTGTCTGCTGCTG 1849
Db 494 -----LeuGlyPheLeuValAsnMetAlaGlySerValValVal 506
QY 1850 TATATGTTCTGCTGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1909
Db 507 ValValIleThrAlaSerAlaValSerLeuIleGlyCysCysPheValAlaLeuPheVal 526

RESULT 3
YHLL1EBV
ID YHLL1EBV STANDARD; PRT; 660 AA.
AC P03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLFI protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrell B.G.;
RT Nature 310:207-211(1984).
RL This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: V01555; -; NOT_ANNOTATED_CDS.
DR F01; A03742; Q08E3.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Alignment Scores:
Pred. No.: 3 46e-11 Length: 660
Score: 398.50 Matches: 237
Percent Similarity: 31.86% Conservative: 37
Best Local Similarity: 27.56% Mismatches: 276
Query Match: 6.40% Indels: 311
DB: 1 Gaps: 51

```

US-09-759-143-110 (1-3410) x YHL1\_EBV (1-660)

```

QY 425 -----CCCCACTTCCAGCAGCAGCAGCGGGGCACATAGGTGATGCTCGCGCCAAACAC 373
Db      ||||| ||| ||| |||||:
542 gGlyHisProProGlyAlaGlyGln-----ArgProSerG1 555
QY 372 ACCTCCAGCCCAAGGTTAGCAGGTTCAGCAGCAGCAGCAGCGTTCGGGTGCGCGACG 313
Db      ||||| ||| ||| |||||:
555 yProThrGlyGlyArg-----ProAlaAlaProGly-----Al 566
QY 312 AGCGGCTCACCACACAGCTCTGGACATAGTGGCCAGCGGGTAGGCTCAGGGGCC 253
Db      ||||| ||| ||| ||||| ||| |||||
566 aProGlyThrProAlaAlaProGlyPro-GlyGlyGlyAlaAlaValProSerGly----- 584
QY 252 GTTCAGGCTCAGCAAGCTCTTCGCTGCTGCTCTCCAGCAGCTCGCGCTCTCTC 193
Db      ||||| ||| ||| ||||| ||| |||||
585 -----AlaThrProHisProGluArg---GlySerGlyPro-----AlaAspProp 599
QY 192 CTTGCTCGCCCACTGCTAGCATCAGCAGCGCGCCCATTCCTGCCAGCCCTTGTGTG 133
Db      ||||| ||| ||| ||||| ||| |||||
599 roAlaAlaAlaArgLeuProGluArg-GlnGluProArgLeuProGluAla 618
QY 132 CGGCTCAGCTCTCAGCCCTGCTCAACACCTGCTGTGGGCGACCTCAGTGGGAC 73
Db      ||||| ||| ||| ||||| ||| |||||
619 AlaAlaGln---ArgCysProAlaGlyProProThrArgSerGlyAlaAlaAlaGln 637
QY 72 AGCTCTCATCTAGATCTCTGGC-----CGAGCGCGCGCTGTCTACCCGGA 25
Db      ||||| ||||| ||||| ||||| ||||| |||||
638 ArgThrHisArgArgProGlyCysProArgSerAlaArg---AsnProGly 654

RESULT 4
CALL_MOUSE
ID CALL_MOUSE STANDARD; PRT; 1453 AA.
AC F11087; 060635;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1 OR COL1A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
of type I procollagen."
RL Matrix Biol. 14:593-595(1995).
RN [2]
RP SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
collagen protein."
RL Gene 39:311-312(1985).
RN [3]
RP SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
evidence for a mouse B1 element within the gene."
RL Mol. Cell. Biol. 2:1362-1371(1982).
RN [4]
RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences."
RL DNA 1:59-69(1981).
RN [5]
RP SEQUENCE OF 1442-1453 FROM N.A.

```

```

RX MEDLINE=88124276; PubMed=3340560;
RA Mooslehner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
of the 3'-untranslated region."
RL Nucleic Acids Res. 16:773-773(1988).
CC -|- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
(FIBRILLAR FORMING COLLAGEN).
CC -|- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -|- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
HYDROXYAPATITE.
CC -|- PM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -|- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U08020; AAA88912.1; -
DR EMBL; X15896; CAA33904.1; -
DR EMBL; M14423; AAA37333.1; -
DR EMBL; M17491; AAA37334.1; -
DR EMBL; X06753; CAA29927.1; -
DR EMBL; K03036; AAA37332.1; -
DR EMBL; K03029; AAA37332.1; JOINED.
DR EMBL; K03030; AAA37332.1; JOINED.
DR EMBL; K03031; AAA37332.1; JOINED.
DR EMBL; K03032; AAA37332.1; JOINED.
DR EMBL; K03033; AAA37332.1; JOINED.
DR EMBL; K03034; AAA37332.1; JOINED.
DR EMBL; K03035; AAA37332.1; JOINED.
DR PIR; A23982; A23982.
DR MGI; 88467; Colla1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib-collagen_C.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Collagen; 1.
DR ProDom; PD002078; Fib-collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWFC; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 29 87 VWFC.
FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 168 1181 TRIPLE-HELICAL REGION.
FT DOMAIN 1182 1207 NONHELICAL REGION (C-TERMINAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
FT CONFLICT 1450 1450 A -> V (IN REF. 5).
SQ SEQUENCE 1453 AA; 137944 MW; 38802E535DF81808 CRC64;

Alignment Scores:
Pred. No.: 4, 31e-10 Length: 1453
Score: 373.00 Matches: 285
Percent Similarity: 31.98% Conservatives: 47
Best Local Similarity: 27.46% Mismatches: 371
Query Watch: 5.99% Indels: 336
DB: 1 Gaps: 55

```

[illegible]



```

QY 792 AGACTGATCATGAAGGCATAGACAGAGTAGGCTGGCCACAGTGGT-----CCGGTGTCC-- 738
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 642 yGluAlaGlyLysProGlyGluGlnGlyValProGlyAspLeuGlyAlaProGlyProse 662
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 737 ---GGAAGAGTCAAGAGCAGAGCCCTCCAGTGGAGTGAAGCACACCTGCCACAGAG 682
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 662 rGlyAlaArgGlyGluArg---GlyPheProGlyGluArgGlyValGlnGlyProGly 681
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 681 TCCAGAGCCGCCACGCCAGGATGAGCAGTGCCA----- 648
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 681 yProAlaGlyProArgGlyAsnAsnGlyAlaProGlyAsnAspGlyAlaLysGlyAspTh 701
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 647 ---GCTCAGGCGCTGGATCCGGCCACAGCAGCCCTGTAGCACCGCCGCTTGGG 592
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 701 rGlyAlaProGlyAla-----ProGlySerGlnGlyAlaProGlyLeuGlnGlyMetPr 719
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 591 ATGAGAAAGAGCTCAGCAGGATGCCACAGGACAGTGGCCAGATGAAGGGCGGGCGG 532
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 719 oGlyGluArgGlyAlaAlaGlyLeuProGly-----ProlysGlyAspArgGlyAspAl 737
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 531 CCATAGCGTCCAGCGCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGCAGGCC 472
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 737 aGlyProLysGlyAlaAspGlySerProGlyLysAspGlyAlaArgGlyLeuThrGlyPr 757
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 471 AGCACTGGACCAATGCCAGCAGCAGCATGTCATGAACCTCTCTCTACCCCACTTCCAGC 412
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 757 oile-----GlyProGlyProAl 764
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 411 AGCAGAGCGGCACATAGTATGCTGCTGGCGCCAAACACACTCTCCAGGC----- 363
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 764 aGlyAlaProGlyAspLysGlyGluAlaGlyProSerGlyProGlyProGlyProThrGlyAl 784
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 362 -----CAAGGTTAGCAGGTTGACCAAGAGAGAGTGGCTTTCGG 322
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 784 aArgGlyAlaProGlyAspArgGlyGluAlaGlyProGlyProAlaGlyPheAlaGl 804
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 321 T---GCCGAGAGCGGCGTCAACACCTCTCCAGCAGCTCT----- 291
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 804 yProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyAspThrGlyVally 824
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 290 -----GGACCATAGTGGCGCAGCGGG---TAGGGCTCAGGGGCGC 253
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 824 sGlyAspAlaGlyProGlyPro-AlaGlyProAlaGlyProGlyProIleGlyAla 844
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 252 GTTCAGGCACTCCAGAACTCTCTGCTGCTGCTGCTCCAGAGTGGCGGCTCTCTC 193
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 844 snValGlyAlaProGly-----ProlysGlyProArgGlyAlaAlaGlyProp 860
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 192 CTTCGCTGCCCACTGCTGCTAGCAATCAGCCAGCGCCCATCTTCTCCAGCCTTGTG 133
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 860 ro-GlyAlaThr-----GlyPheProGlyAlaAlaGlyArgValGlyProGly 876
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 132 CCGGTCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGGCGGCACCTCAGT----- 79
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 877 ProSerGlyAsnAlaGlyProGlyProGlyProGlyProGlyProValGlyLysGlyLys 896
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 78 -----GGGACACGCTCTCATCTCAGTCTCTGGCGCA----- 46
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 897 GlyProArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProGlyPro 916
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 45 -----GGCGGGCGGCTCTCACCAGGAGCC 22
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 917 ProGlyProAlaGlyGluLysGlySerProGlyAla 928
Db      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 5

CAL3\_MOUSE

ID CAL3\_MOUSE STANDARD; PRT: 1464 AA.

AC P08121; Q61429; Q9CRN7;

DT 01-AUG-1988 (Rel. 08, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Collagen alpha 1(III) chain precursor.

GN COL3A1.

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
RC MEDLINE=95011609; PubMed=7926795;
RA Toman D., de Crombrughe B.;
RA "The mouse type-III procollagen-encoding gene: genomic cloning and
RT complete DNA sequence.";
RL Gene 147:161-168(1994).
RN [2]
RP SEQUENCE OF 1-488 FROM N.A.
RC MEDLINE=88167858; PubMed=3443309;
RA Wood L., Thieriault N., Vogeli G.;
RA "Complete nucleotide sequence of the N-terminal domains of the murine
RT alpha-1 type-III collagen chain.";
RL Gene 61:225-230(1987).
RN [3]
RP SEQUENCE OF 1-28 FROM N.A.
RC MEDLINE=85131189; PubMed=3972847;
RA Liao G., Mudryj M., de Crombrughe B.;
RA "Identification of the promoter and first exon of the mouse alpha 1
RT (III) collagen gene.";
RL J. Biol. Chem. 260:3773-3777(1985).
RN [4]
RP SEQUENCE OF 810-1464 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boifelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Lustyich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [5]
RP SEQUENCE OF 1442-1464 FROM N.A.
RC STRAIN=C57BL/6;
RC MEDLINE=91274355; PubMed=2054384;
RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RT collagen mRNAs.";
RL Biochim. Biophys. Acta 1089:241-243(1991).
CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.
CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLINES.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC O-LINKED GLYCAN CONSISTS OF GLC-CAL DISACCHARIDE (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entitles requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC CC EMBL; X52046; CAA36279.1; -  
 CC CC EMBL; M18933; ABA37338.1; -  
 DR DR EMBL; K03037; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AK019448; BAB31724.1; -  
 DR EMBL; X57983; CAA41048.1; -  
 DR PIR; A22287; A22287.  
 DR PIR; A27353; A27353.  
 DR PIR; S16373; S16373.  
 DR MGD; MGI:86453; Col3a1.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; Fib-collagen\_C.  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Pfam; PF01410; COLFI; 1.  
 DR ProDom; PD000007; Collagen; 1.  
 DR ProDom; PD002078; Fib-collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS01208; VWF; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 23  
 FT PROPEP 24 154  
 FT CHAIN 155 1203  
 FT PROPEP 1204 1464  
 FT DOMAIN 31 90  
 FT DOMAIN 155 169  
 FT DOMAIN 170 1195  
 FT DOMAIN 1196 1464  
 FT CARBOHYD 262 262  
 FT MOD\_RES 262 262  
 FT MOD\_RES 283 283  
 FT MOD\_RES 859 859  
 FT MOD\_RES 976 976  
 FT MOD\_RES 1093 1093  
 FT MOD\_RES 1105 1105  
 FT DISULFID 1195 1195  
 FT DISULFID 1196 1196  
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;  
 Alignment Scores:  
 Pred. No.: 1.89e-09 Length: 1464  
 Score: 358.50 Matches: 318  
 Percent Similarity: 29.57% Conservative: 67  
 Best Local Similarity: 24.42% Mismatches: 391  
 Query Match: 5.76% Indels: 527  
 DB: 1 Gaps: 68  
 US-09-759-143-110 (1-3410) x CAL3\_MOUSE (1-1464)  
 QY 3114 GGGCTGAGATGGACAAAGGCTG-----GGAAAC 3085  
 Db 119 GlyArgAsnGlyAspProGlyLeuProGlyGlnProGlyLeuProGlyProGlySer 138  
 QY 3084 CGCACTTGTCTCTGCTCTGCTCCTCAGTAGCTCCAAACAGGTTGTGGAGCTGTGGGAA 3025  
 Db 139 ProGlyIleCys-----GluSerCysProThrGlyGln 150  
 QY 3024 AGTTGGGGG-----TAGGGGAAAGTTGGGGTAGGGGAAAT 2989  
 Db 151 AsnTyrSerProGlnPheAspSerTyrAspValLysSerGlyValGlyGlyMetGlyGly 170  
 QY 2988 TTGGGAGTGCCTTCATCAGCCAGCTCCTAGAGAGAGTAGAGGGGAGT----- 2940  
 Db 171 TyrProGlyProAlaGlyProProGlyPro-ProGlyProProGlySerSerGlyHisPr 190  
 QY 2939 -----GGAAGTGGGGACCAAGCTGGCCCAAGAGAGAGGGGTGGTTAGGAA 2890  
 Db 190 oGlySerProGlySerProGlyTyrGlnGlyProProGlyGluProGlyGlnAlaGly-- 209

2889 GCGTTGAGACCTGAAGCCGCCACCCCTCTACCTTCTTCAACACCCCTAACCTTGGGTAAACA 2830  
 Db 210 -ProAlaGlyProProGlyProGly----- 218  
 QY 2829 GCATTTGGAATTATCATTTGGGATGAGTAGAATTTCAAGGTCTCTGGTTAGGCATTTTG 2770  
 Db 219 -----AlaLeuGlyProAlaGlyProAlaGlyLysAspGlyGluSerGlyLys 234  
 QY 2769 GGGGGCCAGACCCCGAGAGAGATTCTGGCAATGATGACGCCCAATGACCAAGTATCT 2710  
 Db 234 gProGlyArgProGlyGluArgGlyLeu-----ProGlyProProGlyLys 249  
 QY 2709 CAGGGACCTGATTGTTGGGATCCCCACCCCTACCAAAATATTAGACACCAACACAG- 2651  
 Db 249 elysGlyProAla-----GlyMet-ProGlyPheProGlyMetLys---GlyHisArg 266  
 QY 2650 -----AAAGCTAGCA 2641  
 Db 266 LypheAspGlyArgAsnGlyGluLysGlyGluThrGlyAlaProGlyLeuLysGlyGlu 286  
 QY 2640 ATGGATTCCTTCTACTTCTTAAATAAATAAGTTAAATATTAAATGCTGTCTCTG 2581  
 Db 286 snGlyLeuProGlyAsp-----AsnGlyAlaProGlyP 297  
 QY 2580 TGATGCAACAGAGAGACCAAGCCACATCTGTATAAAGTGAAGGGGGTGGAT- 2522  
 Db 297 roMetGlyProArgGlyAlaProGlyGlu-----ArgGlyArgProGlyLeuProG 314  
 QY 2521 -----CAGCAAAAGAGTCTGTGGCTGAGGGGACCTGTTCTTGTGT 2476  
 Db 314 LysAlaGlyAlaArgGlyAsnAspGlyAlaArgly----- 326  
 QY 2475 GTTGCCCTCAGGACTCTTCCCTACAAATAAGTCAATGTTCAATCCATCGAGGAGT 2416  
 Db 326 ----- 326  
 QY 2415 GTTTCATCTAGAACTCCCATCAAGAGCTATACATAAAGAGCTGCAGGTTAAGGGC 2356  
 Db 326 ----- 326  
 QY 2355 TTAGATGGGAACAGTGTGAGTTTATTCAGTCCCAAAACCCCT- 2306  
 Db 327 --SerAspGlyGlnProGly-----ProGlyProProGlyThra 340  
 QY 2305 -----TCTCTAGTGTCTCACTAGGAGGCT 2278  
 Db 340 laGlyPheProGlySerProGlyAlaLysGlyGluValGlyProAlaGlySerProGly- 359  
 QY 2277 AGCTGTTAACTGAGCTGGGTAAATCCA-----CCTGCAGAGTCCCCG 2233  
 Db 360 -----SerAsnGlySerProGlyGlnArgGlyGluProGlyProGlnGlyH 375  
 QY 2232 ATTCCAGTGCATGGAGCCCTTCTGCTCCTCTGTATAGTCCAGACTGAACCCCTTGG 2173  
 Db 375 isAlaGlyAlaGlnGlyProGlyProGlyProGly----- 386  
 QY 2172 AGGCCCTCAGTCAGCAGCCCTCAGACTGGGGAGAG----- 2135  
 Db 387 -----AsnAsnGlySerProGlyGlyLysGlyGluMetGlyProAlaGlyIleProG 404  
 QY 2134 -----AGGAGAGGAGCCGCCAGCC----- 2114  
 Db 404 LysAlaProGlyLeuIleGlyAlaArgGlyProProGlyProAlaGlyThrAsnGlyIleP 424  
 QY 2113 -----CCAGCTGTGAGTACGACCTACGACCTACGACACACAGGT-----GGCA 2074  
 Db 424 roGlyThrArgGlyProSerGlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyA 444  
 QY 2073 GCAGAGAG-----CCACATTACTTTGGCAGCAACA 2044  
 Db 444 laArgGlyGluArgGlyGluAlaGlySerProGlyIleProGlyProLysGlyGluAspG 464  
 QY 2043 GAAACTGGCGCCAGCCCGCAGCCCATGGGCTAAC-----AGGACCGGG 1996

Db 464 lYlYs-AspGlySerProGlyGluPro--GlyAlaAsnGlyLeuProGlyAlaAlaGlyG 483  
 QY 1995 AGCTGGGACCCAGT-----GAGCGAG 1975  
 Db 483 luArgGlyProSerGlyPheArgGlyProAlaGlyProAsnGlyIleProGlyGluLysG 503  
 QY 1974 GCCTGCCA-----CCCCAATGTCTGGGAAGTTTCTACGCTGAGTATTG 1930  
 Db 503 lYProProGlyGluArgGlyGlyProGlyProAlaGlyPro-----ArgGlyValAlaG 521  
 QY 1929 GCCAGTCCCTCTTGCANAATACTACTCTGTAGCAAAAGTAATGCCAGCAGACCCAGG 1870  
 Db 521 lYcLuProGly-ArgAspGlyThr-Pro-----GlyGlyProG 533  
 QY 1869 CTTGGGCGACACATATAGGCAGTGCAGACTGGCTGAGTGGACAAATGGAGGCCATA 1810  
 Db 533 yIleArgGlyMetProGlySerProGlyGlyProGlyAsnAspGlyLysProGlyProPr 553  
 QY 1809 ACAGGGATG-----GGGCCACTGGGACAGCAAGAGGAC----- 1773  
 Db 553 oGlySerGlnGlyGlySerGlyArgProGlyProGlyProSerGlyProArgGlyG 573  
 QY 1772 -----TATCCAGGATGGGAGGTCCAGGCAGATGCCCGGCCCGGAC- 1730  
 Db 573 nProGlyValMetGlyPheProGlyProLysGly-AsnAspGlyAlaProGlyLysAsnG 593  
 QY 1729 -----CACCTGGCTCGGTGGCTCACCCACACACACACACACGTCACGAGACATCA 1681  
 Db 593 lYcLuArgGlyGlyProGlyGlyProGlyLeuPro----- 604  
 QY 1680 CAGGACAGAGCCCGCAGAGCCGGTGGAGTGGAGAGCCACTGCCTCCAGCACCC 1621  
 Db 605 -----GlyProAlaGlyLys-AsnGlyGluThrGlyProGlnGlyProProGlyPro 621  
 QY 1620 ACCTGTCCTATTAGG-----AAGGGA 1600  
 Db 622 ThrGlyProAlaGlyAspLysGlyAspSerGlyProProGlyProGlnGlyLeuGlnGly 641  
 QY 1599 GCTCCAGGCTTAGG----- 1585  
 Db 642 lIleProGlyThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLys 661  
 QY 1584 -----CCTGGCAGGAAGTGGTCATCAGCTGTCTCCTCAGT 1549  
 Db 662 GlyGluValGlyAlaProGlyAlaProGly-GlyLysGlyAspSerGlyAlaProGlyG 681  
 QY 1548 CTAGCACCTCCAGTGTCCCTCGGTATT-----TGGCGAGGAACACCTGC 1504  
 Db 681 uArgGlyProProGlyThrAlaGlyIleProGlyAlaArgGlyGlyAlaGlyProProG 701  
 QY 1503 TTCTCCCGTGTAGAGG-----AGGCCAGTGTGTAGGGCAGGATGTC 1459  
 Db 701 yProGluGlyGlyLysProAlaGlyProProGlyProProGlyAlaSerGlySerPr 721  
 QY 1458 AGGGCTGAGAAGGTGNAACCGGTGAGCGGTGAAGCTGTACACACGCCACACTGTGG 1399  
 Db 721 oGlyLeu-GlnGlyMetProGlyGluArgGly----- 731  
 QY 1398 GACAGGCATGTGGCAGCCGACCCACAGGAAAGTGCACACTGGCCAAATAGACTGCT 1339  
 Db 732 -----GlyProGlySerProGlyProLysGlyGlyGluProGlyGlyAla 748  
 QY 1338 CGAGTCCGAATCGTGTCA-----CCAGCCGGT----- 1311  
 Db 748 laGlyAlaAspGlyValProGlyLysAspGlyProArgGlyProAlaGlyProIleGlyLp 768  
 QY 1310 --CCATGACACAGAGACACAGCAGATGCGGCAGTGCAGGAACACGCCCGAGCTGC- 1254  
 Db 768 roProGlyProAlaGlyGlnProGlyAsp---LysGlyGluGlySerProGlyLeup 787  
 QY 1253 -----CCATCCGAACGCTTCATCATAGTGTCTCCGGGC 1220

Db 787 roGlyIleAlaGlyProArgGlyGlyProGlyGluArgGlyGluHis-----GlyP 804  
 QY 1219 CTCGGTCCCGGCTCAGCTCTGGCAGCCCTGTGTACAGCCCTCGCCACCAAAATCCGT 1160  
 Db 804 roProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluPro----- 820  
 QY 1159 GTAAACAGCGGTGAAGTTCATGAGTGCATCCAGCTGCACAGCTCAGCCACGAAGACCG 1100  
 Db 820 ----- 820  
 QY 1099 GCGCAGGGTCCGGGCATGCGGCAGCAGACAGCTGGTGCAGCCGGGAAGCAGCGGCCA- 1041  
 Db 821 -----GlyAlaLysGlyGluArgGlyAlaProG 830  
 QY 1040 -----GGTTCGGAAAGCAAGCGGCCGCGCATGGACAGCAGTGGGGCGACAAGGA 989  
 Db 830 lYcLuLysGlyGluGlyProProGlyProAla----- 841  
 QY 988 GGGGGCGCAGACCCCTTCTGCTGGCTCGGTGGGGCCAGCGCTGCTCTCAGCCACAG 929  
 Db 842 --GlyProThr-----GlySerSerGlyProAlaGlyProProGlyProGlnG 857  
 QY 928 CAGTGTGGCTGCTACGACGAGGTGAGGAAGATGAGGTGAGCAGGCCAAAGAGGACCTCT- 870  
 Db 857 lYVal-----LysGlyGluArgGlySerProG 866  
 QY 869 -----CCTGGTGGCCAGGTAGGGGCCAGGGCAGTGGTGTCCAGTCAATGCGCAGCAG 815  
 Db 866 lYcLuProGlyThrAlaGlyPheProGlyGlyArgGlyLeuProGlyProGlyAsnA 886  
 QY 814 GAGGTAGCCCGCAGCAGCCCGCAGACTGATCATGAAGGCATACAGAGTAGGCTTGCGG 755  
 Db 886 snGlyAsnProGlyProProGlyPro-----SerGlyAlaProGlyL 900  
 QY 754 ACAGTGTGCTCG-----GGTCCCGGAAGAGGTGCAGAGCAG 719  
 Db 900 yAspGlyProProGlyProAlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaG 920  
 QY 718 GGCCTCCAGTGGAGTGAAGCAGCACCTGGCCACAGAAAGTCCAGCAGCCCGCCAGGAT 659  
 Db 920 lYProLysGlyAlaGlyGlnProGlyGlyLysGlyPro-----ProGly- 935  
 QY 658 GAGCAGTCCCGCAGTCCAGGGGCTGGGATCG-----GSCACAGCAGCCCTCTAGCCAGCC 602  
 Db 936 --AlaGlnGlyProProGlySerProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaA 955  
 QY 601 GGCCTTTGGATGAGAAGAGGCTCAGCAGGATGCCCAAGCAGACAGTCCAGATGAAGGG 542  
 Db 955 rgGlyLeuAlaGlyProProGlyMetProGly---ProArgGlySerProGlyProGlnG 974  
 QY 541 CCGCGCGCGGCATAGC-----GTCCACCGCAGTGTGTACTGGCTGAGGCTAGGAGCGG 488  
 Db 974 lYlIleLysGlyGluSerGlyLysProGlyAlaSerGlyHis-----AsnG 989  
 QY 487 GACACAGA-----CCAGGCCCGCAGCAGTGGACCAATGCCCGCAGCCATGTGTACACT 434  
 Db 989 lYcLuArgGlyProProGlyPro----- 997  
 QY 433 CTCCTCTACCCCACTTCCAGCAGCAGAGGCGGCACATAGGTGATGCTCGCGGCCAACA 374  
 Db 997 lNcGlyLeuProGlyGlnProGlyThrAlaGly-----GluProGlyAla 1011  
 QY 373 CACCTCCAGGCCAAAGTTTAGCAGTTGACACAGAGAGCTGGGCTTTCGGGTGCC----- 318  
 Db 1011 rgAspGlyAsnProGlySerAspGlyGlnProGlyArgAspGlySerProGlyGlyLysG 1031  
 QY 317 -----GCAGCAGGCGGCTCACCCAGCAGCTCTGGACCATAGTGGG--- 278  
 Db 1031 lYAspArgGlyGluAsnGlySerProGlyAlaProGlyAlaProGly-HisProGlyPro 1050  
 QY 277 CCAGCGGGTAGGCGCTCAGGGCGGCTTCCAGGCACTCCAGAACTGTCTCTCGGCTCT 218  
 Db 1051 ProGlyProValGlyProSerGlyLysSerGly-----AspArgGlyGluThrGlyPro 1068



QY 2248 ----CCTGCAGAGTCCCGCATTCAGTCATGGAGCCCTTCTGGCTCCCTGTATAGTC 2192  
DB 109 SerProGlySerProGlyArgThrGlyAlaProGlyProGlyGlnPro-----125  
QY 2191 CAGACTGAACCCCTTGAAGCCCTCCAGTCAGCCAGCCCTAGAGACTGGGAGAGAGG 2132  
DB 126 -----GlySerPro-----128  
QY 2131 AGAGGAGCCCGCCAGCCCGCAGCTGTGACGCTACGCACCTCAGCAGCACAGGTTGGCAGC 2072  
DB 129 ---GlyAlaProGlyProProGlyIleCysGlnSerCysProSerIleSerGlyGlySer 147  
QY 2071 AGAGGCCACATATAC-----TTTGGCAGCAACAGAACTGGCGGCCA 2030  
DB 148 PheSerProGlnTyrAspSerTyrAspValLysAlaGlySerValGlyMetGlyTyrPro 167  
QY 2029 GCCCGGCAG-----CCCATGGGGTACAGAGCGGGGAGCTGGGA---CCC 1985  
DB 168 ProGlnProIleSerGlyPheProGlyProGlyProGlyProSerGlyProGlyProPro 187  
QY 1984 AGTGAGCGAGCCCTCCACCCCAATGTGCTGGAAGTTTTCTACGCTGAGTATTGGCCAA 1925  
DB 188 GlyHisAlaGlyProPro-----GlySer-----195  
QY 1924 GTGCGCTCTTGTCAATACCTACCTGTGTAGCAAGTAATGGCGACACAGAC-----CCAGGC 1869  
DB 196 -----AsnGlyTyrGlnGlyProProGly 203  
QY 1868 CTGCGGCACACACCATATATAGGCAGTGTACAGACTGGCTGAGCTGGACATGGAGCCCATAA 1809  
DB 204 GluProGlyGlnProGlyProSerGlyProProGlyProAlaGlyMetIleGlyPro---222  
QY 1808 ACAGGATGGCGCCACTGGGACAGCAGCAGGAGCCTATCCAGGATGGCGAGTCCAGGC 1749  
DB 223 -----AlaGlyProProGlyLysAspGlyGlu-----ProGly 233  
QY 1748 AGATGCCCGCGCGCGAAC-----CA 1728  
DB 234 Arg---ProGly-ArgAsnGlyAspArgGlyIleProGlyLeuProGlyHisLysGlyH 252  
QY 1727 CCTGCGCTCGGTGGCTCACC-----CACCACCA 1698  
DB 252 sProGlyMetProGlyMetProGlyMetLysGlyAlaArgGlyPheAspGlyLysAspG 272  
QY 1697 CAGGTACGAGACATCACAGGAGAGCCCGCCAG-----ACGCGG 1656  
DB 272 yAlaLysGlyAsp-SerGlyAlaProGlyProLysGlyGluAlaGlyGlnProGlyAla 292  
QY 1655 GTGAGGTGGAGCAGCCACTGCCTCCAGCAGCCCGCTGTCCATTAGGGAAG---GGAG 1599  
DB 292 snGlySerProGlyGlnProGlyProGlyProGlyProThrGlyGluArgGlyArgProGly 312  
QY 1598 CTCACGGTTAGGGCT-----GGCAGAGAGCTGTCTATCAGGCTGCTCCTCAGTGC 1548  
DB 312 snProGly---GlyProGlyAlaHisGlyLysAspGlyAlaProGlyThrAlaGlyPro 331  
QY 1547 TAGCAGCTCCAGTGTCCCTCGGTATTGGGAGGAGAACCTGTCTCCCGGTGGTAGA 1488  
DB 331 euGlyProProGlyProPro-GlyThr---AlaGlyPheProGlySerProGlyPheLys 349  
QY 1487 GGGAGG-----CCAGTGTCTAGGGCAGGATCTGCAGGGCTGAGAGGTGA 1443  
DB 350 GlyGluAlaGlyProProGlyProAlaGlyAlaSerGlyAsnProGlyGluArg-GlyG 369  
QY 1442 ACCCGGTGGGGCTGAGCTGTCTCCACCGCCACACTGTGGCAGGCGATGTGGCAC 1383  
DB 369 uProGlyProGlnGly--GlnAlaGlyProProGlyProGlnGlyProProGlyArgAla 388  
QY 1382 CGGCAGCCAGGGAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGGCGGAATCGCT 1323  
DB 389 GlySerProGlyGlyLysGlyGluMetGlyProSerGlyIle-----402

QY 1322 GCACCAACCCGGTCCATGACCA-----GAG 1299  
DB 403 ---ProGlyGlyProGlyProGlyArgGlyLeuProGlyProGlyProGlyThrSer 421  
QY 1298 AGAAGACACGAGAGTGGCGCACTGCAGGAACACCC-----1263  
DB 422 GlyAsnProGlyAlaLysGlyThrProGlyGluProGlyLysAsnGlyAlaLysGlyAsp 441  
QY 1262 CCAGGCTGCCCATCCGAACGCTTCCATCATAGTGTCTCCGGGCTCGGTGCCCGCTCAG 1203  
DB 442 ProGlyProLysGlyGluArgGlyGluAsnGlyThrProGlyAlaArgGlyProProGly 461  
QY 1202 CTCTGGCAGCCCTGTGTACAGCCCTCCGCCAGAAATCCGTGTAAACAGCGTGAAGG 1143  
DB 462 GluGlyLysArgGlyAlaAsnGlyGluProGly-----Gln 474  
QY 1142 TCATGAGTGCATCCAGCTGCACAGCTCAGCAGAGAGCCCGCGCA-----1095  
DB 475 AsnGlyValProGlyThr-----ProGlyGluArgGlySerProGlyPheArg 490  
QY 1094 -----GGGTGGCGGCGATGGCGCAGCACAGCTGGTGCAGCCGGGAGAGCGCGC 1044  
DB 491 GlyLeuProGlySerAsnGlyLeuProGlyGluLysGlyProAlaGlyGluArgGlySer 510  
QY 1043 CCAGGTCCGGAAGCAAGCCGCGCGATGGACAGCAGTGGCGGACAGAGAGGGG 984  
DB 511 ProGlyProProGlyProSerGlyProAlaGlyAsp-----ArgGly 524  
QY 983 CCGACAGCCCTTCTGCTGGTGGCGCGCGCGCTGCCT-----942  
DB 525 GlnAspGly-----GlyProGlyLeuProGlyMetArgGlyLeuPro 538  
QY 941 -----CCTCAGCAGCAGCA-----GTGTGGCTG 918  
DB 539 GlyIleProGlySerProGlySerAspGlyLysProGlyProGlyAsnGlnGlyGlu 558  
QY 917 CTACGCGAGGTAGGAGATGAGGTGAGCAGCGCCAAAGAGGACTCCT-----870  
DB 559 ProGlyArgSerGlyProGlyProAlaGlyProArgGlyGlnProGlyValMetGly 578  
QY 869 -----CCTGGCTGCCAGGT-----ACGGGCGCCAGG 843  
DB 579 PheProGlyProLysGlyAsnGlyAlaProGlyLysAsnGlyGluArgGlyProGly 598  
QY 842 CACTGGTGTCCAGTCAATGGCAGCAGGA-----GGTACCCAGCAGC 798  
DB 599 GlyProGlyThrProGlyProAlaGlyLysAsnGlyAspValGlyLeuProGlyPro 618  
QY 797 CCCCAGACTGATCATGAAGCATAGCAGTAGGCTGGCAGCAGT-----750  
DB 619 ProGlyProAlaGlyProAlaGlyAspArgGlyGluProGlyProSerGlySerProGly 638  
QY 749 -----GGTCCGGTCCGGAAGAGTCTCAGAGCA-----720  
DB 639 LeuGlnGlyLeuProGlyGlyProGlyProAla---GlyGluAsnGlyLysProGlyGlu 657  
QY 719 -----GGGCTCCAGTGGAGTGAAGC 699  
DB 658 ProGlyProLysGlyAspIleGlyGlyProGlyPheProGlyProLysGlyGluAsnGly 677  
QY 598 ACACCTGCCACAGAAGTCCAGCAGCCCA-----CGCCCA 663  
DB 678 IleProGlyGluArgGlyProGlnGlyProProGlyProThrGlyAlaArgGlyPro 697  
QY 662 GGATGAGCAGTCCAGCT-----CCAGGGGCGC 636  
DB 698 Gly-----ProAlaGlySerGluGlyLysGlyProProGlyProProGlyAla 714  
QY 635 TGGGATCCGCGCACAGCCCTGTCTAGCCAGCGCGCTTGGGATGAGAAAGAGGTCA 576  
DB 715 -----ProGlyGlyThrGlyLeuProGlyLeuGlnGlyMetProGlyGluArgGlyAla 732  
QY 575 GCAGGATGCCCAAGCAGACAGTCCCGCAGTAGAGGCGCGC-----GGCGGCCATAGCGTC 522

```

Db 733 SerGlySerProGly-----ProLysGlyAspLysGlyGluProGlyGlyLysGlyAla 750
QY 521 CAC-----GCACTGTGTCAGCTAGGACGGGACACACACAGGC 474
Db 751 AspGlyLeuProGlyAlaArgGlyGluArgGlyAsnValGlyProLeGlyProProGly 770
QY 473 CCAGCACTGACCAATGCCAGCACCATGTCATGAACCTCTCTCTA---CCCCCACTT 417
Db 771 ProAla-----GlyProProGlyAspLysGlyGluThrGlyProAlaGlyAlaProGly 788
QY 416 CCAGCACACAGGGCGGCACATAGTGTGCTCGGCCAAACACACCTCCAGCCCAAAGG 357
Db 789 ProAlaGlySerArgGly-----GlyProGlyGluArg 799
QY 356 -----TTAGCAGGTTGACACAGACAGCTGGCTTCCGTGCC 318
Db 800 GlyGluGlnGlyLeuProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGlu 819
QY 317 GCAGCAGGC-----GGCTCACCCACA 297
Db 820 ProGlyGlyLysGlyGluArgGlyProProGlyLeuArgGlyGluAlaGlyProProGly 839
QY 296 GCCTCTGCACCATAGTGG-----CCAGCGGGTAGGGCTCAGG 258
Db 840 AlaAlaGlyPro-GlnGlyGlyProGlyAlaProGlyProProGlyProGlnGlyVally 859
QY 257 GGGCGGTTGAGCACTCCAGCAAGTCTCTCTCGCTGCTGCTCCAGAGCTCGCGCTC 198
Db 859 sGlyGluArgGlySerPro-----GlyGlyProGlyAlaGlyAlaGlyPh 873
QY 197 TCCTCTGTCGCGGCACTGCTAGCAATCAGCAGCGCCCATTTCTGCCAGCCCTT 138
Db 873 eProGlyAla-----ArgGlyProProGlyProProGlyAsnAsn----- 886
QY 137 TGTGCGGGTCCAGCTTCTCAGCCCATGCTCAACACTGCTGTGTGGGGCAGCTCAGTG 78
Db 887 -GlyAspArgGlyGluSer-GlyProProGlyValProGlyProProGlyHis---ProG 905
QY 77 GGGACAGCTCTCATCTCAGATCTCGGCGCA----- 46
Db 905 lyProAlaGlyAsnAsnGlyAlaProGlyLysAlaGlyGluArgGlyPheGlnGlyProL 925
QY 45 -----GGCGCGGGCTGTCCACCGGAGCCGCGCTGAGGCTGGTTCCTCC 1
Db 925 euGlyProGlnGlyAlaIleGlySerProGlyAlaSerGlyAlaArgGlyProPro 943

RESULT 7
CAL3_HUMAN
ID CAL3_HUMAN STANDARD; PRT; 1466 AA.
AC P02461; Q15112;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUN-1990 (Rel. 13, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin fibroblast;
RX MEDLINE=89350838; PubMed=2764886;
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA Prockop D.J.;
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT chain of human type III procollagen. Differences in protein structure
RT from type I procollagen and conservation of codon preferences.";
RL Biochem. J. 260:509-516(1989).
RN [2]
RP SEQUENCE OF 149-1225 FROM N.A.
RX MEDLINE=89386015; PubMed=2780304;

```

```

RA Janeczko R.A., Ramirez F.;
RT "Nucleotide and amino acid sequences of the entire human alpha 1
RT (III) collagen.";
RL Nucleic Acids Res. 17:6742-6742(1989).
RN [3]
RP SEQUENCE OF 168-398.
RX MEDLINE=77134724; PubMed=557335;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of cyanogen
RT bromide peptides from the amino-terminal segment of type III collagen
RT of human liver.";
RL Biochemistry 16:1158-1164(1977).
RN [4]
RP REVISIONS.
RA Seyer J.M.;
RL Submitted (DEC-1977) to the PIR data bank.
RN [5]
RP SEQUENCE OF 399-727.
RX MEDLINE=7900343; PubMed=687591;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of five
RT consecutive CNBr peptides from type III collagen of human liver.";
RL Biochemistry 17:3404-3411(1978).
RN [6]
RP SEQUENCE OF 728-964.
RX MEDLINE=80198282; PubMed=6246925;
RA Seyer J.M., Mainardi C., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1
RT (III)-CB5 from type III collagen of human liver.";
RL Biochemistry 19:1583-1589(1980).
RN [7]
RP SEQUENCE OF 950-1466 FROM N.A.
RX MEDLINE=88189827; PubMed=3357782;
RA Mankoo B.S., Dalgleish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:2337-2337(1988).
RN [8]
RP REVISION TO 1184.
RX MEDLINE=89098346; PubMed=3211760;
RA Molyneux K., Dalgleish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:11833-11833(1988).
RN [9]
RP SEQUENCE OF 1065-1466 FROM N.A.
RX MEDLINE=85087944; PubMed=6096827;
RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA Rosenbloom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III
RT procollagen.";
RL Nucleic Acids Res. 12:9383-9394(1984).
RN [10]
RP SEQUENCE OF 965-1200.
RX MEDLINE=81208139; PubMed=7016180;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha
RT 1(III)-CB9 from type III collagen of human liver.";
RL Biochemistry 20:2621-2627(1981).
RN [11]
RP SEQUENCE OF 1176-1466 FROM N.A.
RX MEDLINE=85157600; PubMed=2579949;
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT (III) collagen. Partial characterization of the 3' end region of the
RT gene.";
RL J. Biol. Chem. 260:4357-4363(1985).
RN [12]
RP SEQUENCE OF 1161-1200 FROM N.A.
RX MEDLINE=86187804; PubMed=3754462;
RA Miskulin M., Dalgleish R., Kluge-Beckerman B., Rennard S.I.,
RA Toistoshiev P., Brantly M., Crystal R.G.;
RT "Human type III collagen gene expression is coordinately modulated
RT with the type I collagen genes during fibroblast growth.";
RL Biochemistry 25:1408-1413(1986).

```

VARIANT EDS-IV GLU-1014.  
MEDLINE=92316511; PubMed=1352273;  
Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
Pope F.M.;  
"A single base mutation in the gene for type III collagen (COL3A1)  
converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
syndrome type IV. An unaffected family member is mosaic for the  
mutation.";  
Hum. Genet. 89:414-418(1992).  
[23]  
VARIANT EDS-IV ASP-1050.  
MEDLINE=90037070; PubMed=2808425;  
Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
"Single base mutation in the type III procollagen gene that converts  
the codon for glycine 883 to aspartate in a mild variant of  
Ehlers-Danlos syndrome IV.";  
J. Biol. Chem. 264:19313-19317(1989).  
[24]  
VARIANT EDS-IV VAL-1077.  
MEDLINE=91374480; PubMed=1895316;  
Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,  
Pope F.M.;  
"Characterisation of a glycine to valine substitution at amino acid  
position 910 of the triple helical region of type III collagen in a  
patient with Ehlers-Danlos syndrome type IV.";  
J. Med. Genet. 28:458-463(1991).  
[25]  
VARIANT EDS-IV GLU-1173.  
MEDLINE=93022543; PubMed=1357232;  
Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;  
  
Alignment Scores:  
Pred. No.: 3,87e-09 Length: 1466  
Score: 351.50 Matches: 289  
Percent Similarity: 29.85% Conservative: 57  
Best Local Similarity: 24.94% Mismatches: 391  
Query Match: 5.65% Indels: 424  
DB: 1 Gaps: 62  
  
US-09-759-143-110 (1-3410) x CML3\_HUMAN (1-1466)  
  
Qy 3051 AAACAGGGTGTGGAGCTGTGGGAGAAAGTTGGGGTAGGGGAAGTTGGGGTAGGGGA 2992  
Db 161 LysSerGlyValAlaValGlyGlyLeuAlaGlyTyProGlyProAlaGlyProProGly 180  
Qy 2991 AATTTTGGGCAGTGCCTTCATCAGCCAGCTCCTAGAGAGAGTAGAGGGG----- 2943  
Db 181 -----ProGlyProProGlyThrSerGlyHisProGly 193  
Qy 2942 -AGTGGAGTGGGGGAACAGGCTGGGGCAAGAGAGAGGGTGGTTAGGGGAAGCGGTT 2884  
Db 193 rProGlySerProGlyTrGlnGlyProProGlyGluProGlyGlnAlaGly---ProSe 212  
Qy 2883 GAGACCTTGAGAGCCCAACCCCTTACCTTCCTTCAACACCCCTTAACCTTGGGTAACAGCATTT 2824  
Db 212 rGlyProGlyProProGly----- 219  
Qy 2823 GGAATTTATCATTTGGGATGAGTAGAATTTCCAAGTCTCGGTAGGCATTTTGGGGGGC 2764  
Db 220 -----AlaIleGlyProSerGlyProAlaGlyLysAspGlySerGlyArgProGly 237  
Qy 2763 CAGACCCGAGAGAAAGATTCTCGCAATGATCAGCCCAATGATCAGCCCAATGATCAGGG 2704  
Db 237 yArgProGlyGluArgGlyLeu-----ProGlyProProGlyIleLysGly 252  
Qy 2703 ACCTGATTGTGGGGATCCCCACCCCTACCCAAATATTAGACACCAACACAGA----- 2651  
Db 252 yProAla-----GlyIle-ProGlyPheProGlyMetLys--GlyHisArgGlyPheA 269  
Qy 2650 -----AAAGCTAGCAATGGATTCCCTTCTACTTCTTCTTAATAAATAAG 2608  
Db 269 spGlyArgAnGlyGluSerGlyGluThrGlyAlaProGly-----L 283



QY	2607	TTAATATTAAATGCTGTGTCTCTGATGCAACAGAGGACCAACAGCCACATCC	2548	Db		---ProGly-SerGlnGlyGluSerGlyArgPro-	563
Db	283	eulysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaProGlyProMetGlyProArgG	303	QY	1529	CTCGTATTGGGCAAGGAACACCTCTCTCCCGTGTAGAGGAGGACGCTGTGTAGG	1470
QY	2547	TGATAAAGGTAGAGGGGGTGGATCAGCAAAAGACAGTGTGTGGCTGAGGGACC	2488	Db	564	-----GlyProGlyProSerGlyProArgGlyGlnPro-	575
Db	303	lyAlaProGlyGluArgGlyArgProGlyAlaAlaGlyAlaArgGlyAsn-	322	QY	1469	GCAGGATCTGACGGCTGAGAAGTGAACCGGTGAGGGCGGTGAAGCTGTACACCGG	1410
QY	2487	TGTTCTTGTGTCTGCTCCCTCAGGACTCTTCCCTACAAATAGTCATATGTCAAATC	2428	Db	576	---GlyValMetGlyPhe-	582
Db	322	-----	322	QY	1409	CCACACTGTGGACAGGATGTGGCACCGGACGACAGGAAAGCTCCACACTGGCCA	1350
QY	2427	CCATGGAGAGTGTTCATCTCTAGAACTCCCATGCAAGAGCTACATTAACGAAGCTGC	2368	Db	583	ProLysGlyAsnAspGly-	592
Db	322	-----	322	QY	1349	AATAGACTGTCGAGTGCCTGCGTCCACCGGTCATGACGAGGAGAGACCA	1290
QY	2367	AGGTTAAGGGCTTAGA-----GATGGGAACACAGTGCATGAGTTATTACGTCCCA	2314	Db	593	-----AsnGlyGluArgGlyProGlyGlyProGlyProGlnGlyProGlnGlyProPro	608
Db	323	-----AspGlyAlaArgGlySerAspGlyGlnProGly-	335	QY	1289	GGGAGATGGCGCACTGCAGGA-	1251
QY	2313	AAACCCCTTCTCTAGGTGTCTCAACTAGGAGGCTAGCTGTTAAACCTGAGCCTGGGTA	2254	Db	609	GlyLysAsnGlyGluThrGlyProGlnGlyProGlyProThrGlyProGlyGlyAsp	628
Db	335	roGlyProPro---GlyThrAlaGlyPhePro-	346	QY	1250	TCCGAACGCTTCATCATAGTGTCTCCGGGCTCGGTCCCGGCTCAGCTCTGGCACGC	1191
QY	2253	ATCCACTGCAGAGTCCCGCATTCAGTGCATGGAGCCCT	2200	Db	629	Lys-----GlyAspThrGlyProGlyProGlnGlyLeuGlnGlyLeu	643
Db	346	erProGlyAlaLysGlyGluValGlyProAlaGlySerProGlySerAsnGlyAlaProG	366	QY	1190	CCTGGTACA-	1182
QY	2199	TATAAGTCCAGACTGAACCCCTTGGAGGCTCCAGTCAGGAGCCCTAGAGACTGGG	2140	Db	644	ProGlyThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLysGly	663
Db	366	ly-----GlnArgGlyGluProGlyProGlnG	375	QY	1181	-----GCCCTCGCCACCA	1167
QY	2139	GAGAG-----AGGAGAGGAGGCCCCAGCTGTGCAGCTACGACCTCAGCAG	2086	Db	664	AspAlaGlyAlaProGlyAlaProGlyLysGlyAspAlaGlyAlaProGlyGluArg	683
Db	375	lyHisAlaGlyAlaGlnGlyProProGlyProGlyPhe-	391	QY	1166	AATCGTGTAAACAGCGCTGAAGTCAATGAGTGCAGCTCCAGCTCCAGCTCCAGCAGCA	1107
QY	2085	CACAGGTTGGCAGCAGAGCCACATTTACTTTGGCAGCAACAGAACTGGCGCCAGCCC	2026	Db	684	GlyPro-----ProGlyLeuAlaGlyAlaProGlyLeu	694
Db	391	erProGlyGlyLysGlyGlu-----MetGlyProAlaG	402	QY	1106	AGACCGCGCGAGGTGGGGCATGGCGGACAGCAGCTGTGTGACCGGGGAGAGCAGG	1047
QY	2025	GGCAGCCCCATGGGCTAACAGGAGCGGGAGCTGGACCCAGTGCAGCAGGCCCCCTCCAC	1966	Db	695	ArgGlyGlyAlaGlyProProGlyProGlyLysGlyAlaAlaGlyProGlyProGly	714
Db	402	lyLeProGlyAlaProGlyLeuMetGlyAlaArgGlyPro-----ProGlyProAlaG	420	QY	1046	CGCCAGGTCC	1035
QY	1965	CCCAATGTCTGGAATTTCTACGCTG-----AGTATTGGCCAGTGTCTTGTCAAA	1910	Db	715	ProProGlyAlaAlaGlyThrProGlyLeuGlnGlyMetProGlyGluArgGlyGlyLeu	734
Db	420	lyAlaAsnGlyAlaProGlyLeuArgGlyGlyAlaGlyGluProGlyLysAsn-----	437	QY	1034	-----GGAAGCCAGCGGCGCCGCGCATGGACAGCAGT	1002
QY	1909	TACTACCTGTAGCAAAATAATGGCCAGCAGCCCA-----	1871	Db	735	GlySerProGlyProLysGlyAspLysGlyGluProGlyGlyProGlyAlaAspGlyVal	754
Db	438	-----GlyAlaLysGlyGluProGlyProArgGlyGluArgGlyGluAlaG	453	QY	1001	GGGCGACA-----AGGAGGGGCGCCACAGCCCTTCTGTGCGTGGCGGCCAGCG	948
QY	1870	GCCTGGCGCAGACATATAGGCTGACAGCTGGCTGAGCTGCAACAATGGAGCCCAT	1811	Db	755	ProGlyLysAspGlyProArgGlyProThr-----GlyProIle	767
Db	453	lyLeProGlyValProGlyAlaLysGlyGluAspGlyLysAspGlySerProGlyGluP	473	QY	947	CTGCTCTCTCAGCCACAGCAGTGTGGCTGTAGCAGGTGAGGAAGATGAGGTGAGCA	888
QY	1810	AAACAGGATGGGCACTGGGACAGCAGGA---AGGCATCTATCCAGGATGGCA---	1758	Db	768	GlyProProGlyProAla-----GlyGlnProGlyAspLys	779
Db	473	roGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyAlaProGlyPheArgGlyP	493	QY	887	GGCCAAAGAGCAGTCTCTCTGGTGGTCCAGGT---AGGGGGCGGAGGCGACTGTGTGCC	831
QY	1757	-----GGTCCAGCAGATCCCGCGCGGAGCAGCCCTGGCTCGGTGGCTCACCCAC	1703	Db	780	GlyGluGlyGlyAla---ProGlyLeuProGlyIleAlaGlyPro-----	793
Db	493	roAlaGlyProAsnGlyIleProGlyGlyLysGlyProAlaGlyGluArgGlyAlaProG	513	QY	830	AGTCAATGGCAGGAGGAGTGGCCAGGAGGCGCCCAAGACTGATCATGAAGCATAGA	771
QY	1702	CACCACAGTACGAGAGATCACAGGAGGCGCCGCGAGC-----CGCGGTGGAG	1650	Db	794	-----ArgGlySerProGlyGlu-----	802
Db	513	lyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArgAspGlyValProGlyGlyP	533	QY	770	CAGAGTGGCTGCGGACAGTGTGT---CCGGGT---CCCGAAGAGGTTCAGAGCAGGG	717
QY	1649	GTGGGAGAGCCACTGCTCCAGCAGCCCGCTGTCCATTAGGGAAGGAGCTCCAGGCT	1590	Db	803	ThrGlyProProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluProGly	822
Db	533	roGlyMetArgGlyMetProGlySerProGlyGlyProGlySerAspGlyLysProGly-	552	QY	716	CCTCCAGTGGAGTGAAGCAGCACCTGGCCACAGAAAGTCCAGCAGCCCGCCAGGATGA	657
QY	1589	TAGGCGCTGGCAGGAGTGTGTCATCAGGCTGTCTCTACTGCTAGCAGCTCCAGTGTCCC	1530				

```

Db      823 GlyLysGlyGluArgGlyAlaProGlyGluLysGlyGlyGlyPro---ProGlyVal 841
QY      656 GGAGTCCAGCTCCAGG-----GCTGGGATCCGGGACACAGCAGCCTGTAGCC 606
Db      842 AlaGlyProProGlySerGlyProAlaGlyProGlyProGlyProGlyValLysGly 861
QY      605 AGCGCGCCCTGGGATGAGAAAGAGCTCAGCAGGATGCCCAAGACAGTGCCTCAGATGA 546
Db      862 -----GluArgGlySerProGlyGlyPro----- 869
QY      545 AGGGCGCGCGGCGGCATAGCTCCACGCCAGCTGCTACTGCTGAGCCTAGGAGCGGA 486
Db      870 -----GlyAlaAlaGlyPhe-----ProGlyAlaArg 878
QY      485 CACAGACAGCGCCAGCAGCTGGACCAATGCCAGCAGCAGCTCATGAACCTCTCTCTA 426
Db      879 GlyLeuProGlyPro-----ProGlySerAsnGlyAsnProGly 891
QY      425 CCCCCACATTCAGCAGCAGCAGCGCGCACATAGGTGATGCTCGCGCCAAACACACTCA 366
Db      892 ProProGlyProSerGlySer-----Pro 899
QY      365 GGCCAAAGTTAGCAGCTTGACCAAGCAGAGCTGGGCTTTCGGTGCCCGCACAGCGCGC 306
Db      900 GlyLys-----AspGlyProProGlyProAlaGlyAsnThrGlyAlaPro-----Gly 915
QY      305 TCACCCACAGCCTCTCGACCATAGTGG-----CCAGCGCGGTAGGCTCAGG 258
Db      916 SerProGlyValSerGlyPro-LysGlyAspAlaGlyGlnProGlyGlyLysGlySerPr 935
QY      257 GGCGCGCTTCCAGCAGCTCCAGACTGCTTCTCGCTCGGCTCTGCTCA----- 212
Db      935 oGlyAlaGlnGlyProPro-----GlyAlaProGlyProLeuGlyIi 949
QY      211 -----GAAGCTCGCGCTCTCTCTCTCTCTCTGCTCGCGCAACTGCTAG 171
Db      949 eAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetProGly---ProAr 968
QY      170 GAATCAGCAGCGCGCCATTTCTGCGAGCCCTTTGTCGCGCTGCTCAGTCTCTCAG----- 116
Db      968 gGlySerProGlyProGlnGlyValLysGlyGlySerGlyLysProGlyAlaAsnGlyLe 988
QY      115 -----CCATGCTCAACACTGCTGCTGTGGCGCACCTCA 81
Db      988 uSerGlyGluArgGlyProProGlyProGln-GlyLeuProGlyLeuAlaGly---ThrA 1007
QY      80 GTGGGGAGCAGCTCTCATCTCACTCAGATCCTGGC----- 49
Db      1007 laGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuProGlyArgAspGlyS. 1027

QY      48 -----CGAGCGCGCGGCTGTCACTCCCGGAGCC 22
Db      1027 erProGlyGlyLysGlyAspArgGlyGluAsnGlySerProGlyAla 1042

RESULT 8
CALL CANFA
ID      CALL_CANFA          STANDARD;          PRT; 1460 AA.
AC      Q9X5J7;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      collagen alpha 1(I) chain precursor.
GN      COL1A1.
OS      Canis familiaris (Dog).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX      NCBI_TaxID=9615;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Skin;
RA      Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
RT      "Sequence of normal canine COL1A1 cDNA.";
RL      Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

```

```

CC      -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC      (FIBRILLAR FORMING COLLAGEN).
CC      -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC      -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC      UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC      -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; AF153062; AAD34619.1; -.
DR      InterPro; IPR000087; Collagen.
DR      InterPro; IPR000885; Fib_collagen_C.
DR      InterPro; IPR001007; VWFC_C.
DR      Pfam; PF01391; Collagen; 18.
DR      Pfam; PF01410; COLFI; 1.
DR      ProDom; PD000007; Collagen; 1.
DR      ProDom; PD002078; Fib_collagen_C; 1.
DR      SMART; SM00038; COLFI; 1.
DR      SMART; SM00214; VWFC; 1.
DR      PROSITE; PS01208; VWFC; 1.
KW      Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW      Glycoprotein; Collagen; Signal.
FT      SIGNAL 1 22
FT      PROPEP 23 157
FT      CHAIN 158 1214
FT      PROPEP 1215 1460
FT      DOMAIN 34 92
FT      DOMAIN 158 174
FT      DOMAIN 175 1188
FT      DOMAIN 1189 1214
FT      SITE 741 743
FT      SITE 1089 1091
FT      CARBOHYD 1361 1361
SQ      SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Alignment Scores:
Pred. No.: 4,51e-09 Length: 1460
Score: 350.00 Matches: 277
Percent Similarity: 31.11% Conservative: 50
Best Local Similarity: 26.36% Mismatches: 365
Query Match: 5.62% Indels: 360
DB: 1 Gaps: 56

US-09-759-143-110 (1-3410) x CALL_CANFA (1-1460)
QY      2591 CTGTGCTCTGTGAT---GGCAACAGAAGGACCAAGCCACATCTGTATAAAGGTAA 2535
Db      56 lIeCysValCysAspAsnGlyAsnValLeuCysAspValIleCysAspGlu----- 73
QY      2534 GAGGGGGTGGATCATCAGCAAAAAGACAGTCTGTGGCTGAGGGGACCTGTTCTGTGTG 2475
Db      74 -----ThrLysAsnCysProGlyAlaGlnValProProGly-GluCy 87
QY      2474 TTGCCCCCTCAGGACTCTTCCCTACAAATAAGTTCATATGTTCAATCCCATGGAGAGTG 2415
Db      87 SCYsPro-----ValCys----- 91
QY      2414 TTTCATCTAGAACTCCCATGCAAGAGCTACATTAAACAGCTGACAGTTAAGGGCT 2355
Db      92 -----ProAspGlyGluAlaSerProThrAspGlnGluThrGlyVa 106
QY      2354 T----- 2354
Db      106 iGluGlyProLysGlyAspThrGlyProArgGlyProArgGlyProAlaGlyProProGly 126
QY      2353 -AGAGATGGAAACCCAGGTGACTGAGTTTATTACGCTCCCAAAACCCTTCTCTAGGTGT 2295

```



QY 492 AGCG---GGACACAGA-----CCAGGCCCGACACTGGACCA 460  
 Db 754 YLysAspGlyValArgGlyLeuThrGlyProIleGlyProGlyProAla-----GI 772  
 QY 459 ATGCCCA----- 453  
 Db 772 YAlaProGlyAspGlyGluAlaGlyProSerGlyProAlaGlyProThrGlyAlaAr 792  
 QY 452 ----GCACCATGTGATGAACCTCTCTACCCCACTTCAGCAGCAGCGGCACCA 397  
 Db 792 gGlyAlaProGlyAspArgGlyGluProGlyProGlyProAlaGlyPheAlaGly-- 811  
 QY 396 TAGTGATGCTCGCGCCAAACACACTCCAGGC-----CAAGGTAGCAGG 349  
 Db 812 -----ProProGlyAlaAspGlyGlnProGlyAlaLysGI 823  
 QY 348 TTGACCAGCAGAGCTGGGCTTTCCGGTGGCGCAGCAGCGGCTCACCCACAGCCTCTGG 289  
 Db 823 YGluProGlyAspAlaGlyAlaLysGlyAspAla-----GlyProGlyProAlaGI 841  
 QY 288 ACCATAGTGGG---CCAGCGGGTAGGCTCAGCGGGCGGCTTCAGGCACCTCCAGACTGC 232  
 Db 841 YProThrGlyProGlyProIleGlyAsnValGlyAlaProGly---ProLysGlyA 860  
 QY 231 TTGCTCTCGGCTCTGCTCCAGAGCTCGCGCCTCTCTCTGCTCGCGCCAACTGCCTA 172  
 Db 860 laArgGlySerAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArgVal 879  
 QY 171 GGAATCAGCCAGCGGCCCATTTCTGCAGCCCTTGGTCCGCTTCAGCTTCACGCCCA 112  
 Db 880 Gly-----ProProGlyProSerGlyAsnAlaGlyPro 890  
 QY 111 TGCTCAACACCTCTGCTGTGGGCAC-----CTCAGTGGGACACAGTCT 67  
 Db 891 ProGlyProProGlyProAlaGlyLysGluGlyLysGlyAlaArgGlyGluThrGly 910  
 QY 66 CATCACTCAGATCTCGCGCA-----GCCGCGCG 37  
 Db 911 ProAlaGlyArgProGlyGluValGlyProProGlyProProGlyProAlaGlyGluLys 930  
 QY 36 CTGTCACCGGAGCC 22  
 Db 931 GlySerProGlyAla 935

RESULT 9  
 CA44\_HUMAN  
 ID CA44\_HUMAN STANDARD; PRT: 1690 AA.  
 AC P33420;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 4(IV) chain precursor.  
 GN COL4A4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=95014445; PubMed=7523402;  
 RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reenders S.T.;  
 RT "Complete primary structure of the human type IV collagen alpha 4(IV)  
 chain. Comparison with structure and expression of the other alpha  
 (IV) chains.";  
 RL J. Biol. Chem. 269:26172-26177(1994).  
 RN [2]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RX MEDLINE=98196854; PubMed=9537506;  
 RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
 RA Ninomiya Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
 alpha4(IV) collagen chains are arranged head-to-head on chromosome

RT 2q36.";  
 RL FEBS Lett. 424:11-16(1998).  
 RN [3]  
 RP SEQUENCE OF 1219-1690 FROM N.A.  
 RC TISSUE=Eye;  
 RX MEDLINE=9337374047; PubMed=8365481;  
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;  
 RT "cDNA isolation and partial gene structure of the human alpha 4(IV)  
 collagen chain.";  
 RL FEBS Lett. 330:122-128(1993).  
 RN [4]  
 RP SEQUENCE OF 1407-1507 FROM N.A.  
 RX MEDLINE=93054733; PubMed=1429714;  
 RA Kamagata Y., Mattei M.-G., Ninomiya Y.;  
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
 alpha 4 chain of basement membrane collagen type IV and assignment of  
 the gene to the distal long arm of human chromosome 2.";  
 RL J. Biol. Chem. 267:23753-23758(1992).  
 RN [5]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [6]  
 RP VARIANT AS SER-1201.  
 RX MEDLINE=95078927; PubMed=7987396;  
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,  
 RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,  
 RA Smeets H.J.M., Reenders S.T.;  
 RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)  
 collagen genes in autosomal recessive Alport syndrome.";  
 RL Nat. Genet. 8:77-82(1994).  
 RN [7]  
 RP VARIANT BFH GLU-897.  
 RX MEDLINE=96379660; PubMed=8787673;  
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,  
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;  
 RT "Benign familial hematuria due to mutation of the type IV collagen  
 alpha4 gene.";  
 RL J. Clin. Invest. 98:1114-1118(1996).  
 RN [8]  
 RP VARIANTS AS, AND VARIANTS.  
 RX MEDLINE=99011253; PubMed=9792860;  
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,  
 RA Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;  
 RT "Determination of the genomic structure of the COL4A4 gene and of  
 novel mutations causing autosomal recessive Alport syndrome.";  
 RL Am. J. Hum. Genet. 63:1329-1340(1998).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/  
 CC NIDOGEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,  
 CC COCHLEA, LUNG AND BRAIN.  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -1- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF TYPE II AUTOSOMAL  
 CC RECESSIVE FORM OF ALPORT SYNDROME (AS), AN HEREDITARY



QY 2585 CTCTGTGATGGCAACAGAGACCAACAGGCCAC-----ATCCTGATAAAGGTAA 2535  
Db : : : : :  
QY 401 aGlyMetIleGlyProGlyProGlnGlyPheProGlyLeuProGlyLeuProGlyG1 421  
Db : : : : :  
QY 2534 GAGGGGGTGGATCAGCAAAAGACAGTGTCTGGGCTGAGGGACCTGGTCTTGTGTG 2475  
Db : : : : :  
QY 421 uAlaGlyLeuProGlyArgProAspSerAlaProGly----- 433  
QY 2474 TTGCCCCCTCAGGACTCTTCCCTTACAATAAAGTCATATGTTCAATCCCATGAGAGGTG 2415  
Db : : : : :  
QY 433 ----- 433  
QY 2414 TTTTCATCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTCAGGTTAAGGGCT 2355  
Db : : : : :  
QY 433 ----- 433  
QY 2354 TAGAGATGGGAACAGGTGACTGAGTTTATTACGCTCCCAAAACCCCTTCTCTA---GG 2298  
Db : : : : :  
QY 434 -LysProGlyLysProGly-----SerProGlyLeuProGlyAlaProG1 448  
QY 2297 TGTGTCCTCACTA---GGAGCTAGCTGTTAAACCTGAGCCTGGTGGTAATCCACCTCAGA 2241  
Db : : : : :  
QY 448 yLeuGlnGlyLeuProGlySerValIleTyrCysSerValGlyAsnProGlyProG1 468  
QY 2240 GTCCCCGCATTCAGTGCATGGAGCCCTTCTGCCTCCCTGTATAAGTCCAGACTCAAAAC 2181  
Db : : : : :  
QY 468 nGlyIleGlySlyValGlyProProGlyArgGly-----ProLysGlyGly 486  
QY 2180 CCCCTTGGAAAGCCTCCAGTCAGGCAGCCCTAGAGACTGGGGAGAGAGGAGCGCC 2121  
Db : : : : :  
QY 486 sGlyAsnGluGlyLeuGlyCysAlaCysGluProGlyPro-----MetGlyProPr 502  
QY 2120 CCAGCCCCAGCTGTGAGCTAGCAGCCTCAGCAGCAGCAGGTTGGCAGCAGAGCCACA 2061  
Db : : : : :  
QY 502 oGlyProProGlyLeuProGlyArg---GlnGlySerLysGlyAspLeuGlyLeuProG1 521  
QY 2060 TTACTTTGGAGCAACAAGAACTGGCGCCA-----GCCCGCAGCC 2018  
Db : : : : :  
QY 521 yTrpLeuGlyThrylsGlyAspProGlyProProGlyAlaGluGlyProProGlyLeuPr 541  
QY 2017 -----CATGGGCTAAC----- 2006  
Db : : : : :  
QY 541 oGlyLysHisGlyAlaSerGlyProProGlyAsnLysGlyAlaLysGlyAspMetValVa 561  
QY 2005 -----AGGAGCGGGAGCTGGGACCCAGTGGAGCGGCCCTCCA----- 1967  
Db : : : : :  
QY 561 lSerArgValLysGlyHisLysGlyGluArgGlyProAsp-----GlyProProGlyPh 579  
QY 1966 -CCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAAGTCGCTTCTCAATA 1908  
Db : : : : :  
QY 579 eProGlyGlnProGlySerHisGlyArg-----AspGlyHisAlaGly----- 593  
QY 1907 CTACCTGTAGCAAAATAATGCGACCA----- 1876  
Db : : : : :  
QY 594 -----GluLysGlyAspProGlyProProGlyAspHisGluAspAlaTh 608  
QY 1875 CCCAGGCTCGGCAGACACATATAGCAGTGCAGACTGCTGAGCTGGACATGAG 1816  
Db : : : : :  
QY 608 rProGlyGlyLysPheProGlyProLeuGlyProProGlyLysAlaGlyProValG1 628  
QY 1815 CCCATAACAGGATGGGGCCACTCTGGGACAGCAGGAAGGCAC----- 1773  
Db : : : : :  
QY 628 yPro---ProGlyLeuGlyPheProGlyProProGlyGluArgGlyHisProGlyValPr 647  
QY 1772 -----TATCCAGGATGGGAGTCCAGCAGATGCCCGGCCGCGAA----- 1731  
Db : : : : :  
QY 647 oGlyHisProGlyValArgGlyProAspGlyLeuLysGlyGlnLysGlyAspThrIleSe 667  
QY 1730 -----CCACCTGGCC 1720  
Db : : : : :  
QY 667 rCysAsnValThrTyrProGlyArgHisGlyProProGlyPheAspGlyProProGlyPr 687  
QY 1719 TCGGTG-----GG 1712

Db : : : : :  
QY 687 oLysGlyPheProGlyProGlnGlyAlaProGlyLeuSerGlySerAspGlyHisLysG1 707  
QY 1711 CTCACCCACCACACACACATAGGAGACATCACAGCAGAGCCCGGAGAGCGGGGT-- 1654  
Db : : : : :  
QY 707 yArgProGlyThrProGlyThr-----AlaGluIleProGlyProProGlyPh 723  
QY 1653 -----GGAGGT-----GGAGCAGGCCACTG---CC 1631  
Db : : : : :  
QY 723 eArgGlyAspMetGlyAspProGlyPheGlyGlyLysSerProValGlyPr 743  
QY 1630 TCCACAGCCACCGTCTCCATTAGGC-----AAGGAGCTCCAGC----- 1591  
Db : : : : :  
QY 743 oProGlyProProGlySerProGlyValAsnGlyGlnLysGlyIleProGlyAspProAl 763  
QY 1590 -----TTAGGG-----CCTGCAGAGAAGCT 1571  
Db : : : : :  
QY 763 aPheGlyHisLeuGlyProProGlyLysArgGlyLeuSerGlyValProGlyIleLys-G 783  
QY 1570 GGTCACTAGGCTGCTCACTCACTAGCAGCTCCAGTGTCCCTCGGTATT----- 1521  
Db : : : : :  
QY 783 yProArgGlyAspProGlyCysProGlyAlaGluGlyProAlaGlyIleProGlyPheL 803  
QY 1521 ----- 1521  
QY 803 euGlyLeuLysGlyProLysGlyArgGluGlyHisAlaGlyPheProGlyValProGlyP 823  
QY 1521 ----- 1521  
QY 823 roProGlyHisSerCysGluArgGlyAlaProGlyIleProGlyGlnProGlyLeuProG 843  
QY 1520 ---TGGCAGGAACACCTGCTCTCCCGGTGTAGAGGAGGCCA-----GTGTGTAGGG 1469  
Db : : : : :  
QY 843 lyTyProGlySerProGlyAlaProGlyLysGlyGlnProGlyAspValGlyProp 863  
QY 1468 CAGGATCTCAGGCTGAGAGGTGAACCCGCTGAGGCGCGCTGAAGCTGTCAACACGCC 1409  
Db : : : : :  
QY 863 roGlyProAlaGlyMetLys-----GlyLeuProGlyLeuProGlyArgP 878  
QY 1408 CACACTGTGGGACAGCATGTGGCAGCCAGCCAGCCAGGAAAGCTGCCACACTGGCCAA 1349  
Db : : : : :  
QY 878 roGlyAla-----HisGlyProPro-----GlyLeuProGlyIleProG 891  
QY 1348 ATAGACTGTCTGAGTCCGCAATCGCTGCACACCGCGTCCA---TGACAGAGAGAAGAC 1292  
Db : : : : :  
QY 891 lyProPheGlyAspAspGlyLeuProGlyProProGlyProLysGlyProArgGlyLeuP 911  
QY 1291 CAGGAGATGGCGCACTCAGGAACAGCCCGCAGGCTGCCATCCCAAGCCCTTCATCATA 1232  
Db : : : : :  
QY 911 roGly-----PheProGlyPheProGlyGluArgGlyLysProG 924  
QY 1231 GTGTCTCCGGGCTCGGTCCCGGCTCAGCTGTGGCAGCCCTGGTACAGCCCTCGCC 1172  
QY 924 lyAlaGluGly-----CysProGlyAlaLys---GlyGluProGlyGluLysGlyMetS 941  
QY 1171 CACGAATCCGTGTAAACACAGCAGTGAAGTCATGAGTGCCATCCAGCTCCACAGCTCAGC 1112  
Db : : : : :  
QY 941 eGlyLeuPro-----GlyA 946  
QY 1111 CACAGAGCCCGCGCAGGGTCGGGGGCAATGCGGC----- 1077  
Db : : : : :  
QY 946 sPArgGlyLeuArgGlyAlaLysGlyAlaLysGlyProProGlyAspGluGlyGluMetA 966  
QY 1076 -----AGCACAGTGGTCCAGCCGGGAGAGAGCGGCCCGCCAGGTTC----- 1035  
Db : : : : :  
QY 966 laIleSerGlnLysGlyThrProGlyGluProGlyProProGlyAspGlyPheP 986  
QY 1034 -----GAAAAGCCAGCGCGCGGCTGACAGCAGTGGGCGCACAGGAGGG 986  
Db : : : : :  
QY 986 roGlyGluArgGlyAspLysGlyThrProGlyMetGlnGly-----ArgArg 1002  
QY 985 GCGCAGACACCCCTTCTGCTGGCTCGGTG----- 957  
Db : : : : :

Db 1002 lyGlu-----LeuGlyArgTyrGlyProGlyPheHisArgGlyGluProG 1018  
 QY 956 -----GGCCAGCGTCCCTCTCAGCCACAGCAGTGTGCTGTAC 914  
 Db 1018 lyGluLysGlyGlnProGlyProGlyProGlyProGlySerThrGlyLeuA 1038  
 QY 913 GCAGG---TGAGGAAGATGAGGTGAGCAGGCCAAAGAGGCACTCCCT-----CCTGGGT 863  
 Db 1038 rgGlyPheIleGlyPheProGlyLeuProGlyAspGlnGlyGluProGlySerProGlyP 1058  
 QY 862 GCCAGGT-----AGGGGGCCA----- 846  
 Db 1058 roProGlyPheSerGlyIleAspGlyAlaArgGlyProLysGlyAsnLysGlyAspProA 1078  
 QY 845 --GGGCACTGG----- 827  
 Db 1078 laSerHisPheGlyProGlyProLysGlyGluProGlySerProGlyCysProGlyH 1098  
 QY 826 AATGGCAGGAGGAGGTAGCCAGCCAGCCCAAGACTGATCATGAAGCATAGACAGA 767  
 Db 1098 lsPheGlyAlaSerGlyGlnGlyLeuProGlyIleGlnGlyProArgGlySerProG 1118  
 QY 766 GTAGGCTGGCAGCTGGT-----CCGGGT-----C 740  
 Db 1118 lyArgProGlyProGlySerSerGlyProGlyProGlyCysProGlyAspHisGlyMetP 1138  
 QY 739 CCGGA---AGAGGTCAGAGCAGGCGCTCCAGTGGAGTGAACACACCTGGCCACAGA- 684  
 Db 1138 roGlyLeuArgGlyGln-----ProGlyGluMetGlyAspProGlyProArgG 1154  
 QY 683 -----AGTCCAGCAGCCCA---CGCCAGGATGAGCAGTGCCTA----- 648  
 Db 1154 lyLeuGlnGlyAspProGlyIleProGlyProGlyIleLysGlyProSerGlySerP 1174  
 QY 647 -----GCTCCAGGGG- 638  
 Db 1174 roGlyLeuAsnGlyLeuHisGlyLeuLysGlyGlnLysGlyThrLysGlyAlaSerGlyL 1194  
 QY 637 CCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCGCTGGGATGAGAAGAGGCT 578  
 Db 1194 euHisAspValGlyProGlyProValGlyIleProGlyLeuLysGlyGluArgGlyA 1214  
 QY 577 CAGCAGATGCCAAGACAGT-----CCAGATGAAGCGCGCGCGGCCATA 527  
 Db 1214 spProGlySerProGlyIleSerProGlyProArgGlyLysGlyProGlyP 1234  
 QY 526 GCGTCCAGCGCAGTGGTCACTGGTGCAGCCAGCAGCGGACACAGA----- 480  
 Db 1234 roProGlySerSerGlyProGlyProAlaGlyAlaThrGlyArgAlaProLysAspI 1254  
 QY 479 -----CCAGGCCAGCAGTGCACCAATGCCCA----- 453  
 Db 1254 leProAspProGlyProGlyAspGlnGlyProGlyProAspGlyProArgGlyA 1274  
 QY 452 -----GCACCATGTGTCAT 440  
 Db 1274 laProGlyProProGlyLeuProGlySerValAspLeuLeuArgGlyGluProGlyAspC 1294  
 QY 439 GAATCTCTCTACCCCACTTCCAGCAGCAGCGGCGGCATAGTGTGCTCGCGC 380  
 Db 1294 ysGlyLeuProGlyProGlyProGlyProGly-----P 1305  
 QY 379 CAACACACCTCCAGGCCAAAG-----TTAGCAGGTGTGACACAG 341  
 Db 1305 roProGlyProGlyTyrLysGlyPheProGlyCysAspGlyLysAspGly-GlnLys 1324  
 QY 340 CAGAGCTGGCTTCGGTGGCGCAG---CAGGCGGCTCACCACAGCCTCTGGACCAT 284  
 Db 1325 GlyProMetGlyPheProGlyProGlnGlyProHisGlyPheProGlyProGlyGlu 1344  
 QY 283 AGTGGG---CCAGCGGGTAGGCTCAGGGGGCGCTCAGGCACTCCAACTGCTTCGT 227  
 Db 1345 LysGlyLeuProGlyProGlyArgGlyLysGlyProThrGlyLeuPro----- 1360

QY 226 CTGGCTGCTCTCCAGAGCTGGGCTCTCTCTCTGTC-----TGCCGCCAA 179  
 Db 1361 -----GlyProArgGlyGluProGlyProProAlaAspValAspCysProArg 1377  
 QY 178 CTCCTAGGAATCAGCCAGCGGCCCATTTCTGCCACCCCTTTGGTCCCGTCCAGCTTCT 119  
 Db 1378 IlePro-GlyLeuProGlyAlaProGlyMetArgGlyPro----- 1390  
 QY 118 CAGCCCATGCTCAACACCTGCTGCTGTGGGCGCACCTCAGTGGGACACGCTCTCATCCTC 59  
 Db 1391 -----GluGlyAlaMetGly---LeuProGly-----MetArgG 1401  
 QY 58 AGATCTGCGCGAGGCGCGCTGTACCCGAGCAGCGCGTGCAGG 10  
 Db 1401 yProProGlyProGlyCysLysGlyGluProGlyLeuAspGlyArg 1417

RESULT 10  
 ID CAL3\_BOVIN STANDARD; PRT: 1049 AA.  
 AC P04258;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(III) chain.  
 GN COL3A1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE OF 1-242.  
 RX MEDLINE=80026026; PubMed=488906;  
 RA Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wachter E.,  
 Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. I. The amino  
 acid sequence of the amino terminal region of the alpha 1(III) chain  
 (positions 1-222).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).  
 RN [2]  
 RP SEQUENCE OF 243-422.  
 RX MEDLINE=80026027; PubMed=488907;  
 RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. II. The amino  
 acid sequence of the cyanogen bromide peptide alpha 1(III)CBI,8,10,2  
 (positions 223-402).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).  
 RN [3]  
 RP SEQUENCE OF 423-571.  
 RX MEDLINE=80026028; PubMed=488908;  
 RA Bentz H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. III. The  
 amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4  
 (positions 403-551).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).  
 RN [4]  
 RP SEQUENCE OF 572-808.  
 RX MEDLINE=80026029; PubMed=488909;  
 RA Lang H., Glangville R.W., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. IV. The amino  
 acid sequence of the cyanogen bromide peptide alpha 1(III)CB5  
 (positions 552-788).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).  
 RN [5]  
 RP SEQUENCE OF 809-947.  
 RX MEDLINE=80026030; PubMed=488910;  
 RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. V. The amino  
 acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A  
 (position 789-927).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).  
 RN [6]



SEQUENCE OF 948-1049.  
 RX MEDLINE=80026031; PubMed=488911;  
 RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. VI. The amino  
 acid sequence of the carboxyterminal cyanogen bromide peptide alpha  
 1(III)CB9B (positions 928-1028).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).  
 CC 1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 CC 1- ALONG WITH TYPE I COLLAGEN.  
 CC 1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 CC ALSO CROSS-LINKED VIA HYDROXYLINES.  
 CC 1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC PIR: A02862; CGB07S.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR001007; WVF\_C.  
 DR Pfam: PF01391; Collagen; 17.  
 DR ProDom: PD000007; Collagen; 1.  
 DR ProSITE: PS01208; WVFC; PARTIAL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen.  
 FT DOMAIN 1 14  
 FT DOMAIN 15 1040  
 FT DOMAIN 1041 1049  
 FT MOD\_RES 95 95  
 FT MOD\_RES 107 107  
 FT MOD\_RES 119 119  
 FT MOD\_RES 938 938  
 FT MOD\_RES 950 950  
 FT CARBOHYD 107 107  
 FT CARBOHYD 950 950  
 FT DISULFID 1040 1040  
 FT DISULFID 1041 1041  
 FT SEQUENCE 1049 AA; 93651 MW; 8EEC33D1C66EC9A3 CRC64;  
 Alignment Scores:  
 Pred. No.: 8.19e-09 Length: 1049  
 Score: 344.50 Matches: 298  
 Percent Similarity: 30.79% Conservative: 53  
 Best Local Similarity: 26.14% Mismatches: 405  
 Query Match: 5.53% Indels: 385  
 DB: 1 Gaps: 64  
 US-09-759-143-110 (1-3410) x CAL3\_BOVIN (1-1049)  
 QY 3051 AAACAGSGTGTGGAGCTGGTGGGAAAGTTGGGGTAGGGAGGGA 2992  
 DB 8 LysSerGlyValAlaGlyGlyGlyIleAlaGlyTyrProGlyProAlaGlyProGly 27  
 QY 2991 AATTTTGGGAGTGCCTTCATCAGCCCTCCTCAGAGAGAGTAGAGGG- 2943  
 DB 28 -----ProGlyPro-ProGlyThrSerGlyHisProGlyAl 40  
 QY 2942 -ACTGGAAGTGGGGGAACAGCAGCTGGGCGCAGAGAGAGGGTGTAGGGAAGCCGTT 2884  
 DB 40 aProGlyAlaProGlyTyrGlnGlyProGlyGluProGlyGlnAlaGly---ProAl 59  
 QY 2883 GAGACCTGAAGCCCAACCTCTACCTTCTTCAACACCCCTACCTTGGGTAACAGCATTT 2824  
 DB 59 aGlyProGlyProProGly----- 66  
 QY 2823 GGAATTATCTTTGGGTAGT-----AGAAATTCAGGTCTCT 2785  
 DB 67 -----AlaIleGlyProSerGlyLysAspGlyGluSerGlyArgProGly-ArgProG 84  
 QY 2784 GGGTAGGCATTTGGGGGCCAGACCCAGGAGAGAGAGATCTGGCAATGATCAGCCC 2725  
 DB 84 lyProArgGlyPheProGlyPro----- 91  
 QY 2724 AATGACCATCTATCTCAGGGACCTGATTGTTGGGATCCCCACCCCTACCCAAATATTA 2665  
 DB 92 -----ProGlyMetLysGlyPro-----AlaGlyMetProGlyPheProGlyMetL 107

2664 GACACCAACACAGA-----AAAGCTAGCAATGATTCCTT 2629  
 QY : : : : :  
 DB 107 ys--GlyHisArgPheAspGlyArgAsnGlyGluLysGlyGluProGlyAlaProG 126  
 QY 2628 CTACTTTGTTAAATAAATAGTTAAATATTAATGCTGTCTGTCTGTATGGGCAACAG 2569  
 DB : : : : :  
 DB 126 ly-----LeuLysGlyGluAsnGlyValProGlyGluAspGlyAlaP 140  
 QY 2568 AAGGACCAACAGCCACATCTGATAAAGCTAAGAGGGGGTGGATCAGCAAAAGACA 2509  
 DB : : : : :  
 DB 140 roGlyProMetGlyProArgGlyAlaProGlyGluArgGlyArgProGlyLeuProGlyA 160  
 QY 2508 GTGCTGTGGCTCAGGGACCTGGTCTTCTGTGTGGTGGCCCTCAGGACTCTCCCTTACA 2449  
 DB : : : : :  
 DB 160 laAlaGlyAlaArgGlyAsn----- 166  
 QY 2448 AATAAGTCATATGTTCAATCCCATCGAGGAGTGTTCATCTCTAGAACTCCCATCGAAG 2389  
 DB : : : : :  
 DB 166 ----- 166  
 QY 2388 AGCTACATTAAACAGAGCTGAGGTTAAGGGCTTAGA-----GATGGAAACACAGGTG 2335  
 DB : : : : :  
 DB 167 -----AspGlyAlaArgGlySerAspGlyGlnProGly- 177  
 QY 2334 ACTGAGTTTATTCAGCTCCCAAAACCTTCTCTAGTGTGTCTCACTAGGAGGTAGC 2275  
 DB : : : : :  
 DB 178 -----ProGlyProPro---GlyThrAlaGlyPhePro----- 188  
 QY 2274 TGTTAACCTGAGCCTGGTAATCCACCTCAGAGTCCCGCATTCAGTGATGGAGCC 2215  
 DB : : : : :  
 DB 189 -----GlySerProGlyAlaLysGlyGluValGlyProAlaGlySerP 203  
 QY 2214 CTTCTGGCCTCCTGTATATAAGTCCAGACTGAAACCCCTTTGGAAGGCTCCTCAGTCAAGCA 2155  
 DB : : : : :  
 DB 203 roGly-----SerSerGlyA 208  
 QY 2154 GCCCTAGAGACTGGGAGAG-----AGGAGAGGAGCCGCCAGCCGCCAG 2110  
 DB : : : : :  
 DB 208 laProGlyGlnArgGlyGluProGlyProGlnGlyHisAlaGlyAlaProGlyProPro- 227  
 QY 2109 CTGTGCAGCTACGACCTCAGCAGCAGAGGTTGGCAGCAGAGAGCCACATCTATTGTCGA 2050  
 DB : : : : :  
 DB 228 -----GlyProGlySerAspGlySerPro-----GlyG 238  
 QY 2049 GCAACAGAACTGGCGCCAGCCCGCAGCCCATCGGCTAACAGAGCGGGAGGTGG 1990  
 DB : : : : :  
 DB 238 lyLysGlyGluMetGlyProAlaGlyIleProGlyAlaProGlyLeuIleGlyAlaArg 258  
 QY 1989 GACCCAGTGGCAGGAGCCCTCCACCCCATGCTGGAAGTTTCTACGTAGTATTG 1930  
 DB : : : : :  
 DB 258 lyPro-----ProGlyProProGlyThrAsnGlyValProGlyGlnArgGlyAlaAla 276  
 QY 1929 GCCAAGTCGCTTGTCAATACTACCTGTGTAGCAAAAGTAATGGCAGCAGACCCAGG 1870  
 DB : : : : :  
 DB 276 lyGluProGlyLys-----AsnGlyAlaLysGlyAspProGlyProA 290  
 QY 1869 CCTCGGCAGACACATATAGGAGTGCAGACTGGCTGAGCTGAGTGCAGCAATGGAGCCCAT 1810  
 DB : : : : :  
 DB 290 rgGlyGluArgGlyGluAlaGlySerProGlyIleAlaGlyProLysGlyGluAspGlyL 310  
 QY 1809 AACAGGATG-----GGCCACCTGGCAGCAGCAGGA---AGGC 1775  
 DB : : : : :  
 DB 310 ysAspGlySerProGlyGluProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArg 330  
 QY 1774 ACTATCCAGGATGGCAGGTCAGGATGC-----CCCGGCCGGAA---CCAC 1727  
 DB : : : : :  
 DB 330 lyValProGlyPheArgGlyProAlaGlyAlaAsnGlyLeuProGlyGluLysGlyProp 350  
 QY 1726 CTGCGCTCGTGGGCTCACCACACACAGTACGAGACATCAGCAGCAGGAGGCC 1667  
 DB : : : : :  
 DB 350 roGlyAspArgGlyGlyProGlyProAlaGlyProArg-GlyValAlaGly-GluProG 369

QY 1666 GCAGAGCCG-----GGTGGAGTGGAGCAGGCCACTGCTCCAGCAGCCAGCCAGTGTCT 1613  
 Db : : : : :  
 369 yArgAsnGlyLeuProGlyGlyProGlyLeuArgGlyLeuArgGlySerProGlyGlyPr 389  
 QY 1612 ATTAGGAGGAGGAGCTCCAGCTTAGGCTTGGAGGAGGAGTGTGTCTAGGCTGTCTC 1553  
 Db : : : : :  
 389 oglySerAsnGlyLysProGly---ProProGly-SerGlnGlyGlyThrGlyArgPro- 407  
 QY 1552 ACTGTAGCAGCTCCAGTGTCCCTCGTATTGGGAGGAGAACACCTGCTCTCTCCGGTG 1493  
 Db : : : : :  
 408 : : : : :  
 QY 1492 GPAGAGGAGGCCAGTGTGT----- 1473  
 Db : : : : :  
 415 roArgGlyGlnProGlyValMetGlyPheProGlyProLysGlyAsnAspGlyAlaProG 435  
 QY 1472 --AGGCGAGGATCTCCAGGCTGAGAGGTGAACCCGTGAGGCGGTGAAGTGTCTAC 1415  
 Db : : : : :  
 435 lyLysAsnGlyGlu-ArgGlyGlyProGlyProGlyProGlnGlyProAlaGlyLys 454  
 QY 1414 CACGGCCACACT-----GTGGACAGGCGATGTGGCAGGCGGACAGGAGGAGTGTG 1361  
 Db : : : : :  
 455 AsnGlyGluThrGlyProGlnGlyProGlyProThrGlyProSerGlyAspLysGly 474  
 QY 1360 CACACTGGCCAAATAGACTGTCTGAGTCCGAATGCTGCACCA-----GCCGGTCCAT 1307  
 Db : : : : :  
 475 AspThrGlyProProGlyProGlnGlyLeuGlnGlyLeu---ProGlyThrSerGlyProp 494  
 QY 1306 GACACAGAGAGA-----AGACAGGAGAGTGGCGACTGCTGAGGAGGAGGAGGAG 1259  
 Db : : : : :  
 494 roGlyGlyAsnGlyLysProGlyGluProGlyProLysGlyGlyAlaGly---AlaProG 513  
 QY 1258 GCTGCCATCCAGCGCTTATCATAGTGTCTCGGCGCTCGGTGCGGCGCTCAGTCT 1199  
 Db : : : : :  
 513 lyLeuProGlyGlyLysGlyAspSerGlyAlaProGlyGlyLeuArgGlyPro----- 529  
 QY 1198 GGGCAGCGCTGGTACAGCC-----CCTCGCCACAGAAATCCGTGTAAACAGCGTGA 1145  
 Db : : : : :  
 530 : : : : :  
 QY 1144 GGTATGAGTCCAGCTGACAGCTGACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGG 1085  
 Db : : : : :  
 541 : : : : :  
 QY 1084 CATGCGGC-----AGCAGAGTGTGTCAGCGGAGGAGGAGGAGGAGGAGGAGGAG 1040  
 Db : : : : :  
 553 laAlaGlyProProGlyProProGlyProGlyProGlyProGlyProGlyProGlyPro 573  
 QY 1039 GTTCC---GGAAAGCCAGCGGCGCGG-----ATGGACAGCA 1004  
 Db : : : : :  
 573 lyGluArgGlyGlyProGlyGlyProGlyProLysGlyAspLysGlyGluProGlySers 593  
 QY 1003 GTGGGGGAGCA-----AGAGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 965  
 Db : : : : :  
 593 erGlyValAspGlyAlaProGlyLysAspGlyProArgGlyProThr----- 608  
 QY 964 CTCGTGGGGCGGCGGCTGCTCTCAGCCAGCAGGAGTGTGGTGTCTACGAGGTGAG 905  
 Db : : : : :  
 609 : : : : :  
 QY 904 GAAGATGAGGTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 848  
 Db : : : : :  
 618 lyGlnProGlyAspLysGlyGluSerGlyAla---ProGlyValProGlyLeuAlaGlyP 637  
 QY 847 CAGGCGACTGTGTCTCCAGTCAATGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 788  
 Db : : : : :  
 637 ro-----ArgGlyGlyProGlyGlu----- 643  
 QY 787 GATCATGAGGATAGACAGAGTAGGCTGGCGAGAGTGTGT---CCGGGT---CCCGGAA 734  
 Db : : : : :  
 644 : : : : :  
 QY 733 GAGGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 674

Db : : : : :  
 661 InAsnGlyGluProGlyAlaLysGlyGluArgGlyAlaProGlyGlyGluGlyGlyGly 681  
 QY 673 CCCCAGCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAG 623  
 Db : : : : :  
 681 lyPro---ProGlyAlaAlaGlyProAlaGlyGlySerGlyProAlaGlyProGlyP 700  
 QY 622 CAGCAGCCCTGTAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 563  
 Db : : : : :  
 700 roGlnGlyValLysGly-----GluArgGlySerProGlyGlyProG 714  
 QY 562 GGACAGTCCAGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 509  
 Db : : : : :  
 714 lyAla-----AlaGlyPheProGlyGlyArgGly----- 723  
 QY 508 ACTGGGTGACCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 449  
 Db : : : : :  
 724 : : : : :  
 QY 448 CATGGTTCATGAACCTTCTCTCTACCCCA-----CTTCAGCAGCAGGAGGAGGAG 401  
 Db : : : : :  
 728 roGlySerAsnGlyAsnProGlyProProGlySerSerGlyAlaProGlyLysAspGly- 747  
 QY 400 CACATAGGTGATGCTGCGGCGCAACACACCTCCAGGCGCAAGAGTGTAGCAGGTTGAC 341  
 Db : : : : :  
 748 : : : : :  
 QY 340 CAAGAGTGGGCTTTCGGTCCGCGCAGCAGCAGC-----GGCTCAC 302  
 Db : : : : :  
 759 lySerProGlyLysSerGlyProLysGlyAspSerGlyProProGlyGluArgGlyAlaP 779  
 QY 301 CCAGAGCTCTGAGCAGTGTGGG-----CCAGGCGGAGTGTGGGCGGCGGCGGCGGCGG 245  
 Db : : : : :  
 779 roGlyProGlnGlyPro-ProGlyAlaProGlyProLysGlyLeuAlaGlyLeuThrGly 798  
 QY 244 ACTCAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 194  
 Db : : : : :  
 799 AlaArgGlyLeuAlaGlyProProGlyMetProGlyAlaArgGlySerProGlyProGln 818  
 QY 193 : : : : :  
 819 GlyLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 838  
 QY 163 CCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 116  
 Db : : : : :  
 839 ProGly-----ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGlu 853  
 QY 115 CCCATGCTC-----AACACCTGCTGTGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62  
 Db : : : : :  
 854 ProGlyArgAspGlyAsn-ProGlySerAspGlyLeuProGlyArgAspGlyAlaProG 873  
 QY 61 CTCAGATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 22  
 Db : : : : :  
 873 yAlaLysGlyAspArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 886

RESULT 11  
 CALL\_CHK  
 ID CALL\_CHK STANDARD; PRT: 1453 AA.  
 AC P02457;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Collagen alpha 1(I) chain precursor.  
 GN COL1A1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE OF 1-153 FROM N.A.  
 RX MEDLINE=88056316; PubMed=3678834;  
 RA Finer M.H., Boedtker H., Doty P.;



Db 182 Pro-----GlyAlaProGlyProGlnGlyPheGln----- 191  
 QY 2092 TCAGCAGCACAGGCTGGCAGACAGCCACATTACTTTGGCAGCACAGAACTGGCGG 2033  
 Db 192 --GlyProProGlyGluProGlyGluProGlyAlaSerGlyProMetGlyProArgGly 210  
 QY 2032 CCAGCCGGCAGCCCACTGGGCTAACAGGAGGGGAGCTGGGA-----CCC 1985  
 Db 211 ProAlaGlyProProGlyLysAsnGlyAspGlyGluAlaGlyLysProGlyArgPro 230  
 QY 1984 AGTCAGCAGCCCTCCA-----CCCCAATGTCTGGAAG- 1950  
 Db 231 GlyGlnArgGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 250  
 QY 1949 -----TTTCTACGCTGAGTATTGGCCAAAGCTCTTGT 1914  
 Db 251 ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAsp----- 263  
 QY 1913 CAATATACCTGTGTAGCAAGTAAGTAAAGCCAGACCA-----GGCCTGGCGGACACA 1857  
 Db 264 -----GlyAlaLysGlyGlnProGlyProAlaGlyProLysGlyGlu 277  
 QY 1856 CCATATAGCAGTCACAGACTGCTGAGCTGGACATGAGGCCCAATAACAGGATGGGG 1797  
 Db 278 ProGlySerProGlyGluAsnGlyAlaProGlyGlnMetGlyPro-----ArgGly 294  
 QY 1796 CCACCTGGGACAGCAGGAGCACTATCCA-----GGATGGCAGGTCCAGGC 1749  
 Db 295 LeuProGlyGluArgGlyArgProGlyProSerGlyProAlaGlyAlaArgGlyAsnAsp 314  
 QY 1748 AGATGCCCCGGCGGAA-----CCACCTGGCTCGTGGTGGCTCACCCACCAACA----- 1696  
 Db 315 GlyAlaProGlyAlaAlaGlyProProGlyProThrGlyProAlaGlyProProGlyPhe 334  
 QY 1695 -----CCTACGGAGACATCACAGCAGAGGCCCGCCGAGAGCGCGGT 1654  
 Db 335 ProGlyAlaAlaGlyAlaLysGlyGluThrGlyProGlnGlyAlaArgGlySerGlyGly 354  
 QY 1653 GGAGGTGGGAGCAGG-----CCACTGCTCCAGCAGCCACCGTGTCCATTAGG-----AAG 1603  
 Db 355 ProGlnGlySerArgGlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAla 374  
 QY 1602 GGAGCTCCAGCTTAGGG-----CCTGGCAGGAAGCTGTCTATCAGGTGCTCTCCTACTG 1549  
 Db 375 GlyAsnProGlyAlaAspGlyGlnProGlyAlaLys-GlyAlaThrGlyAlaPro----- 392  
 QY 1548 CTAGCACTCCAGTCCCTCGTATTTGGCAGGAACACTGCTCTCCCGGTGGTAG 1489  
 Db 393 -----GlyIle---AlaGlyAlaProGlyPheProGlyAlaAr 404  
 QY 1488 AGGG-----AGGCCAGTGTGTAGGCAGGATCTGCAGGCTGAGAGGTGAACCGGNG 1435  
 Db 404 gGlyProSerGlyProGlnGlyProSerGlyAlaProGlyProLys-GlyAsnSerGlyG 424  
 QY 1434 AGGGCGGTGAAGCTGTACCAGCGCCACACT----- 1403  
 Db 424 luProGlyAlaProGlyAsnLysGlyAspThrGlyAlaLysGlyGluProGlyProAlaG 444  
 QY 1402 -----GTGGACAGCATGTGCGCAGCGCCAGCCCGCCAGGAGAA 1366  
 Db 444 lyValGlnGlyProProGlyProAlaGlyGluGlyLysArgGlyAlaArgGlyGluP 464  
 QY 1365 GCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGA----- 1329  
 Db 464 roGlyProAlaGlyLeuProGlyProAlaGlyGluArgGlyAlaProGlySerArgGlyP 484  
 QY 1328 -----ATCGTGTACACCGCGCTCCATGACCAGAGAGA----- 1296  
 Db 484 heProGlyAlaAspGlyIleAlaGlyProLysGlyProProGlyGluArgGlySerProG 504  
 QY 1295 -----AGACAGGGAGATGGCGACTG 1274  
 Db 504 lyAlaValGlyProLysGlySerProGlyGluAlaGlyArgProGlyGlyLeuP 524

QY 1273 CAGGAACA-----GCCCAAGCTGCCCATCCGACAGCCCTTCATCATATAGTG 1229  
 Db 524 roGlyAlaLysGlyLeuThrGlySerProGlySerProGlyPro-----AspGlyL 541  
 QY 1228 TCTCCGGGCTCGGTCCCGGCTCAGCTCTGGGCGAGCCCTGTACAGCCCTCGCCAC 1169  
 Db 541 ystThrGlyProProGlyProAlaGlyGlnAspGlyArgProGlyProAlaGlyProPro 560  
 QY 1168 GAAATCCGTGTAAACAGCGTGAAGGTCTAGTGCCATCCAGCTGCACAGCTCAGCCAC 1109  
 Db 561 -----GlyA 562  
 QY 1108 GAAGACCCCGCAGGCTCGGGGATGCGGCAGACACACTGGTCAGCGCGGGAGAGCAG 1049  
 Db 562 laArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyGluProG 582  
 QY 1048 GCGCCCGAGTTCC-----GGAAAGCCAGCGCGGCGCATGACAGCAGTGGGGGACAA 992  
 Db 582 lyLysProGlyGluArgGlyAlaProGlyProProGly-----AlaValGlyAlaAlaG 600  
 QY 991 GGAGGGGCGCAGACGCCCTTCTGCTGGCTCGGTGGGCGCCAGCGCTGCTCTCAGCCAC 932  
 Db 600 lyLysAspGlyGluAla-----GlyAlaGlnGlyProProGlyProT 614  
 QY 931 CAGCAGTGTGCTGTACCCAGCTGAGGAAGTAGGGTGACAGCGGCAAGAGAGCACT- 873  
 Db 614 hr-----GlyProAlaGlyGluArgGlyGluG 623  
 QY 872 -----CCTCTGGTGGTCCAGGT---AGGGG---CCAGGGCAGTGGTGTCCAGTCAAT 824  
 Db 623 lnglyProAlaGlyAlaProGlyPheGlnGlyLeuProGlyProAlaGlyProPro----- 641  
 QY 823 GCGAGCAGGAGGTAGCCCGCAGCAGCCCGCCAGACTGATCATGAAGCATAGACAGSTA 764  
 Db 642 -----GlyGluAlaGlyLysProGlyGlu-----GlnGlyV 652  
 QY 763 GGCCTGGCGCAGTGGT---CGGGTCCC-----GGAGAGGTGAGAGAGGCGGCTC 713  
 Db 652 alProGlyAsnAlaGlyAlaProGlyProAlaGlyAlaArgGlyGluArg---GlyPheP 671  
 QY 712 CAGTGGAGTGAAGCAGACACTGGCCACAGAAAGTCCAGCAGCCCGCCAGCAGGATGAGCAG 653  
 Db 671 roGlyGluArgGlyValGlnGlyProProGlyProGlnGlyProArgGlyAlaAsnGlyA 691  
 QY 652 TGCCA-----GCTCCAGGGGCTGGATCCGGGCA 623  
 Db 691 laProGlyAsnAspGlyAlaLysGlyAspAlaGlyAlaProGlyAla-----ProGlyA 709  
 QY 622 CAGCAGCCCTGCTAGCCAGCGCGCTTGGATGAGAAAGAGGTCTCAGCAGGATGCCCA- 564  
 Db 709 snGluGlyProProGlyLeuGlyMetProGlyGluArgGlyAlaAlaGlyLeuProG 729  
 QY 563 -----AGGACAGTGGCCAGATGAAGGGCGCGCGGCGC----- 531  
 Db 729 lyAlaLysGlyAspArgGlyAspProGlyProLysGlyAlaAspGlyAlaProGlyLysA 749  
 QY 530 -----CATAGCGTCCACCCAGTGGTC 509  
 Db 749 spGlyLeuArgGlyLeuThrGlyProIleGlyProProGlyProAlaGlyAlaProGly- 768  
 QY 508 ACTGCTCAGCCTAGGAGCGGACACAGCAGCCAGCCCGCCAGCA-----CTGACCAANTGCC 455  
 Db 769 -----AspLysGlyGluAlaGlyProProGlyProAlaGlyProThrGlyAlaArgG 786  
 QY 454 CAGCACCATGTGTATGAATCTCTCTACCCCGCTTCCAGCAGCAGCAGCGCGCACATA 395  
 Db 786 lyAlaProGlyAspArgGlyGluProGlyProProGlyProIleGlyPheAlaGly----- 804  
 QY 394 GGTGATGCTCGGCGCAACACACTCCAGGCCAAAGGTTAGCAGGTTGACCAAGAG 335  
 Db 805 -----ProProGly-----AlaAspGlyGlnProGlyAlaL 815





Qy	1679	AG-----GCAGAGCCCGCCGACAGCCGGGTGGAGGTGGAGCAGGACCCAC	1635
Db	357	roGlnGlyProArgGlySerGluGlyProGlnGlyValArgGly-----GluProG	374
Qy	1634	TGCCTCAGCAGCCACCGTGTCCATTAGG---AAGCGAGCTCCAGGCTTAGGG-----C	1584
Db	374	lyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnP	394
Qy	1583	CTGGCAGGAAGCTGGTTCATCAGGCTGTCCTACTGCTAGCACCTCCAGTGTCCCTCGGT	1524
Db	394	roGlyAlaLys-GlyAlaAsnGlyAlaPro-----Gly	404
Qy	1523	ATTGGCGAGGAACACTGCTTCTCCGGTGGTAGAGGGAGGCCACTGTGTAGGGCAGGA	1464
Db	405	Ile---AlaGlyAlaProGlyPheProGlyAlaArgGly-----ProSerGly	419
Qy	1463	TCATGCAGGCTGAGAGGTTCAACCGGTGAGGCGGCTGAAGCTGTCCACCACGGCCACAC	1404
Db	420	Pro-GlnGlyProGlyProGlyProGlyProLysGlyAsnSerGlyGluProGly---Al	438
Qy	1403	TGTGGGACAGGCATGTGGCACCGCAGCCACAGGAAAGCTGCACACTGGCCAAATAGA	1344
Db	438	aproGlySerLysGlyAspThrGly-AlaLysGlyGluProGlyPro-----	453
Qy	1343	CTGCTCAGTGGCCGAATCGTGTGCACAGCCGGTCCATGACCAGAG-----AGAAGACCA	1290
Db	454	-----Val-GlyValGlnGlyProProGlyProAlaGlyGluGlnGlyLysArgGly	470
Qy	1289	GGGAGATGGCGACTGCAGGAACAGCC---CCAGGCTGCCATCCGAACGC-----	1242
Db	471	AlaArgGlyGluProGlyProThrGlyLeuProGlyProGlyProGlyAlaArgGlyGlyPro	490
Qy	1241	-----CTTCATCATAGTGTCTCCGGGCGTCCGGTCCCGCTCAGCTC	1200
Db	491	GlySerArgGlyPheProGlyAlaAspGlyValaGlyProLysGlyProAlaGlyGlu	510
Qy	1199	TGGCAGCCCTGTGTACAGCCCTCGCCACGAAATCCGTGTAAACAGCGCTGAAGTCA	1140
Db	511	ArgGlySerProGlyProAlaGlyProLysGlySerPro-----	523
Qy	1139	TGAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGAGCCGGCGGCGGATGC	1080
Db	523	-----	523
Qy	1079	GGCAGCACAGCTGTGTGCAGCCGGGGAAGCAGGCGCCACAGT-----TCCGGA	1032
Db	524	---GlyGluAlaGlyArgProGlyGlyAlaGlyLeuProGlyAlaLysGlyLeuThrGly	542
Qy	1031	AAGCCAGCGGGCCCGCATGGACAGCAGTGGGGCGACAGGAGGGGGCCGACAGCCCTT	972
Db	543	SerProGlySerProGlyProAspGly-----LysThrGlyProPro-----	556
Qy	971	CTGCTGCTCGCTGGGGCCCGCAGC-----CTGCTCTCTCAGCCACCA	930
Db	557	-----GlyProAlaGlyGlnAspGlyArgProGlyProProGlyProPro	571
Qy	929	-----GCAGTGTGGCTG	918
Db	572	GlyAlaArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAla-----	589
Qy	917	CTAGCCAGGTGAGGAAGATCAGGTGACAGGCGCAAGAGGCACACT-----CTCTCTGGG	864
Db	590	-----GlyGluProGlyLysAlaGlyGluArgGlyValProGlyProProGly	605
Qy	863	TGCCCCAGGT-----AGG	852
Db	606	AlaValGlyProAlaGlyLysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAla	625
Qy	851	GGGCCAGGGCCTGGGTCTCCAGTCAATGCAGGCAGGAA-----	813
Db	626	GlyProAlaGlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeu	645

RESULT 13	
CA15_HUMAN	
ID	CA15_HUMAN
AC	P20908;
DT	01-FEB-1991 (Rel. 17, Created)
	STANDARD;
	PRT; 1838 AA.







QY 1802 ATGGGCGACCTGGG---ACAGCAGGAGGACCATATCCAGGATGCGCAGGTCCAGGCAGA 1746  
 Db 768 GluGlyProProGlyGluLysGlyGlyGlnGlyProProGlyProGlnGlyProGlnGly 787  
 QY 1745 TGCCCGCGCCGGAA----- 1731  
 Db 788 TyrProGlyProArgGlyValLysGlyAlaAspGlyLeuArgGlyLeuLysGlyThrLys 807  
 QY 1731 ----- 1731  
 Db 808 GlyGluLysGlyGluAspGlyPheProGlyPheLysGlyAspMetGlyLeuLysGlyAsp 827  
 QY 1730 ----- 1731  
 Db 828 ArgGlyGluLeuGlyProGlyProArg---GlyGlu----- 839  
 QY 1685 CATCAGCAGCAGGCGCGGAGCGGGTGGAGGTGGGAGCAGGCGCACTGCCCTCCAG 1626  
 Db 840 --AspGlyProGlyProLysGlyArgGlyGlyProAsnGlyAspProGlyProLeu 859  
 QY 1625 CACCCACGCTCCATTAGGGAAG---GGAGCTCCAGGCTTAGGG-----CCTGCGCAGA 1575  
 Db 859 lypProGlyGluLysGlyLysLeuGlyValProGlyLeuProGlyTyrProGlyArg 879  
 QY 1574 AGCTGCTCATCAGGCTGCTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATT----- 1521  
 Db 879 ln-----GlyProLysGlySer-IleGlyPheProGly 889  
 QY 1520 TGGGAGGAGAACCTGCTTCTCCCGTGTAGAGGAGGCGCAGTGTGTAGGCGCAGATCT 1461  
 Db 890 PheProGlyAlaAsnGlyGluLysGlyGlyArgGlyThrPro-----GlyLys 905  
 QY 1460 CGAGGCTGAGAAGTGAACCGGTGAGGCGGCTGAAGCTGTACACAGGCGCACACTGT 1401  
 Db 906 ProGly----- 907  
 QY 1400 GGCAGAGGATGTGCACCGCGCAGCCACAGGAAAGCTGCCACACTGGCCAAATAGACTG 1341  
 Db 908 -----ProArgGlyGln----- 911  
 QY 1340 CTCGAGTCCGATCGCTGCACCGCGCTCATGACACAGAGAGA----- 1296  
 Db 912 -----ArgGlyProThrGlyProArgGlyGluArgGlyProArgGlyLeu 926  
 QY 1295 -----AGACACAGG---AGATGGCGCACTGCAGGACAGCCCA---GGCTGCGCATCC 1248  
 Db 927 ThrGlyLysProGlyProLysGlyAsnSerGlyGlyAspGlyProAlaGlyProProGly 946  
 QY 1247 GAACGCTTCATCATAGTGTCTCCGGGCTCGGTGCCGCTGCTGCTGCGCACGCGCT 1188  
 Db 947 GluArg-----GlyProAsnGlyProGlnGlyProThrGlyPhePro 960  
 QY 1187 GGTACAGCCCTCGCCCAAGAAATCCGTGTAACACAGCGGTCAAGTCAATGATGCCA--- 1131  
 Db 961 GlyProLysGlyProProGlyProProGlyLysAspGly-----LeuProGly 976  
 QY 1130 -----TCCAGCTGCACAGCTCAGCCAGCAGAGACC 1101  
 Db 977 HisProGlyGlnArgGlyGluThrGlyPheGlnGlyLysThrGlyProProGlyProPro 996  
 QY 1100 GCGCAGGTCGCGGCGCATCGCGCAGCACA-----GCTGTTGCGAGCGGGGAGCAGG 1047  
 Db 997 GlyVal---ValGlyProGlnGlyProThrGlyGluThrGlyProMetGlyGluArgGly 1015  
 QY 1046 CGCCGAGTTCGCGAAAGCAGCGG-----CCCGCATGGACAGCAGTGGG 999  
 Db 1016 HisProGlyProProGlyProProGlyGlnGlnGlyLeuProGlyLeuAlaGlyLysGlu 1035  
 QY 998 GCGACAGAGGCGGCGCAGACGCTTCTGCTGCTGCGGTGGGCGCCAGCG---CTGCTCT 942  
 Db 1036 GlyThrLysGlyAspPro-----GlyProAlaGlyLeuPro 1047

QY 941 COTCAG-----CCACCAGCAGTGTGGCTGCTACCGAGGTGAGGAAGATGAGGTGAGCA 888  
 Db 1048 GlyLysAspGlyProPro-----GlyLeuArgGlyPhePro 1059  
 QY 887 GCCCAAAAGAGGCACT-----CCT 870  
 Db 1060 GlyAspArgGlyLeuProGlyProValGlyAlaLeuGlyLeuLysGlyAsnGluGlyPro 1079  
 QY 869 CTGGTGTCCAGGT-----AGGGGGCAGGCGCAGTGTGCTCCAGTCAATGCGCAGCA 816  
 Db 1080 ProGlyProProGlyProAlaGlySerProGlyGluArgGlyProAla-----GlyAla 1097  
 QY 815 GGAGTACCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 756  
 Db 1098 AlaGlyProIleGlyIlePro-----GlyArgProGly 1108  
 QY 755 GACAGTGTCCGGTCCCGGGAAGAGGTGAGAGAGCAGGCGCTCCAGTGGAGTGAAGCACA 696  
 Db 1109 ProGlnGlyPro-----ProGlyProAlaGlyGluLysGlyAla 1121  
 QY 695 CTGGCCACAGAAAGTCCAGCAGCGCCAGGATGAGCAGTGTGCGCAGTGTCCAGGCGCC 636  
 Db 1122 ProGlyGluLysGlyProGlnGlyPro-AlaGlyArgAsp-----GlyLeuGlnGlyPr 1139  
 QY 635 TGGATCGGCGCAGCAGCGCTCTGCTAGCCAGCGCGCTTGGGATGAG----- 587  
 Db 1139 o---ValGlyLeuProGlyProAlaGlyProValGlyProProGlyGluAspGlyAsp 1158  
 QY 586 -----AAAGAGTCTCAGCAGGATGCGCCAGGACAGTGTGCGCAGTGTCCAGGCGCC 540  
 Db 1158 sGlyGluIleGlyGlnProGlyGlnLysGlySerLysGlyAspLysGlyGlnGlnGlyPr 1178  
 QY 539 GCGCGCGCCATCGCTCCAGCGCGTGTGCTGCTGAGCTAGCTAGCA-----GCG 489  
 Db 1178 oProGly-ProThrGlyPro---GlnGlyProIleGlyGlnPro-GlyProSerGlyAla 1196  
 QY 488 GGCACAGCAGCG 459  
 Db 1197 AspGlyGluProGlyProArgGlyGlnGlnGlyLeuPheGlyGlnLysGlyAspGluGly 1216  
 QY 458 -----TGCCAGCAGCATGCTCATGAATCT-----CCTCTACCCCA 420  
 Db 1217 ProArgGlyPheProGlyProProGlyProValGlyLeuGlnGlyLeuProGlyProPro 1236  
 QY 419 CTCCAGCAGCAGGCGGCGCACATAGTGTGCTGCGGCCAACACACCTCCAGGCGCAA 360  
 Db 1237 GlyGluLysGlyGluThrGlyAspValGlyGlnMetGlyProGlyProProGlyPro 1256  
 QY 359 AGG---TTAGCAGGTTGACCCAGCAGAGCTGGGCTTCCGGTCCCGCAGCAGCG---GGC 306  
 Db 1257 ArgGlyProSerGlyAlaProGlyAlaAspGlyProGlnGlyProGlyGlyIleGly 1276  
 QY 305 TCACCCAGCAGCTCTGACCATAGTGGG-----CCAGCGGCTAGGCGCTCAGG 258  
 Db 1277 AsnProGlyAlaValGly-GluLysGlyGluProGlyGluAlaGlyGluProGlyProse 1296  
 QY 257 GCGCGCTTACGCACTCCA-----GAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 213  
 Db 1296 rGlyArgSerGlyProProGlyProLysGlyGluArgGlyGluLysGlyGluSerGlyPr 1316  
 QY 212 AGAAGCTGCGGCT 1336  
 Db 1316 oSerGlyAlaAlaGlyProGlyProLysGlyProGlyProGlyProGlyProGlyProGly 1336  
 QY 167 TCAGCAGGCGCGCATTTCTGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120  
 Db 1336 ySerProGly-----ProValGlyPheProGlyAspProGlyProProGly 1351  
 QY 119 TCAGCCCATGCTCAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60  
 Db 1351 yGluPro-----Gly-ProAlaGlyGlnAspGlyProProGlyProGlyProGlyProGly 1369  
 QY 59 CAGATCTGCG 25





Qy 932 -----CCAGCAGTGTGGTCTACGAGGTGAGGAAGATGAG 896  
 Db 1092 lyAlaThrGlyProGlyLeuProGlyLeuGlyGluArg-----GlyTyThrG 1110  
 Qy 895 GGTGAGAGGCCAAAGAGGCACTCT-----CTGGGTGCG 860  
 Db 1110 lySerAlaGlyGluGlyGluProGlyProGlySerGlyGluLeuProGlyProP 1130  
 Qy 859 CAGGT-----AGGGGGCCAGGCGACTGGTCCAGTCAGTGC 821  
 Db 1130 roGlyProAlaGlyProArgGlyGluArgGlyPro----- 1141  
 Qy 820 AGGCAGGAGGTAGCCAGGAGCCCAAGACATGATCATGAAGGCATAGACAGTAGGC 761  
 Db 1142 -----GlnGlyAsnSerGlyGluLysGlyAsp-----GlnGlyPheGlnGly 1157  
 Qy 760 CTGGCGACAGTGTCCGGT-----CCCGGAAGAGGTTCAGAGCAGGCG 716  
 Db 1157 roGlyPheThrGlyProThrGlySerProGlyPheProGlyLysValGlySerProGlyP 1177  
 Qy 715 CTCCA-----GTGAGGTGAAGACACCTGCCACAGAGTCCAGCAGCCCGCCAGCGC 665  
 Db 1177 roProGlyProGlnAlaGlySerGlyGluGlyLeuArgGlyProSerGly--LeuP 1196  
 Qy 664 CAGGATCAGCAGTCCAGCTCCAGGGGCTGGGATCGGGGCACAGCAGCCCTCTAGCCA 605  
 Db 1196 roGlySerProGlyProGlyProGlyProGlyGlyGlnGly-----Pro 1210  
 Qy 604 GCCGCG-----CCTCGGATGAGAAAGAGGCTCAGCAGGATCCCAAGAGCAGTCCCA 551  
 Db 1211 AlaGlyLeuAspGlyLeuAspGlyLysPheProGlyLeuArgGly----- 1227  
 Qy 550 GATGAAGGGCGGGCGG-----GCCATAGCGTCCAGCGGCTGCTGCTGCTGAGCC 497  
 Db 1228 AspProGlyProAlaGlyProProGlyLeuMetGlyProGlyPheLysGly--LysT 1247  
 Qy 496 TAGAGCGGGACACAGACAGCAGCGCCAGCTGACCACTGACCACTCCAGCAGCAGTGTCTAGAA 437  
 Db 1247 hrGlyHisProGlyLeuProGlyProGlyLysAsp--CysGlyLysProGly----- 1263  
 Qy 436 CTCTCTCTCTACCCCA-----CTTCAGCAGCAGCAGCGCG----- 402  
 Db 1264 -----ProGlySerThrGlyArgProGlyAlaGluGlyGluProGlyAlaM 1280  
 Qy 401 -----GCACATAGGTGATCGCTGCGCGCCAAACACAC 371  
 Db 1280 etGlyProGlnArgProGlyProGlyHis-----ValGlyProGlyP 1297  
 Qy 370 CTCAGCGCCAAAGTTAGCAGGTGACACAGCAGCTGGCTTCCGGT----- 321  
 Db 1297 roProGlyGln-----ProGlyProAlaGlyIleSerAlaValGlyLeuL 1312  
 Qy 320 -----GCCGAGCAGGCGGCTCACCCACAG 296  
 Db 1312 ysGlyAspArgGlyAlaThrGlyGluArgGlyLeuAlaGlyLeuProGlyGlnProGlyP 1332  
 Qy 295 CCTGTGGACCATAGTGG-----CCAGCGGGTAGGCTCAGGGGCGGTTC 248  
 Db 1332 roProGly-HisProGlyProProGlyGluProGlyThrAspGlyAlaAlaGlyLysGlu 1351  
 Qy 247 GGCACCTCAGAACTGCTCGCTCGCTCTCCAGAGCTGGCGCTCTCTCTCTCTTC 188  
 Db 1352 GlyProGlyLysGlnGlyPheTyThrGlyProGlyProGlyProGlySerGlyAspProGlyAla 1371  
 Qy 187 TCCGCGCAACTGCTAGGAATCAGCAGCGCCCATTTCTGCCAGCCCTTGTGGCGGT 128  
 Db 1372 AlaGlyGln-----LysGlyGlnAlaGlyLysGlyArgAlaGlyMetProGlyGly 1389  
 Qy 127 CCAGCTTCTCAGCCCATCTCAACACCTGCTGCTGCTGGGACCTCAGTGGGACACGTC 68  
 Db 1390 ProGlyLysSerGlySerMetGly--ProValGlyProGlyProAlaGlyGluArgL 1409  
 Qy 67 TCATCACTCAGATCTGCGCCGAGGCGCGGCTGTCACCCGGA 25

Db 1409 yHisProGlyAlaProGlyProSerGly-----SerProGly 1421  
 RESULT 15  
 STP\_SPIOL STANDARD; PRT; 525 AA.  
 AC Q03411;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sucrose transport protein (Sucrose permease) (Sucrose-proton symporter).  
 OS Spinacia oleracea (Spinach).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
 OC NCBI\_TaxID=3562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Leaf;  
 RX MEDLINE=93099843; PubMed=1464305;  
 RA Riesmeier J.W., Willmitzer L., Frommer W.B.;  
 RT "Isolation and characterization of a sucrose carrier cDNA from spinach by functional expression in yeast";  
 RT EMBO J. 11:4705-4713(1992).  
 RL  
 CC CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).  
 CC -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF SUCROSE INTO THE CAN ALSO TRANSPORT MALTOSE AT A LESSER RATE.  
 CC -!- PATHWAY: Sucrose metabolism.  
 CC -!- SUBCELLULAR LOCATION: Inner membrane.  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X67125; CAA47604.1;  
 CC PIR; S28052; S28052.  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr; 1.  
 DR TIGRFAMs: TIGR01301; GPH\_sucrose; 1.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 KW Transmembrane; Transport; Sugar transport; Symport.  
 FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 38 58 1 (POTENTIAL).  
 FT TRANSMEM 72 92 2 (POTENTIAL).  
 FT TRANSMEM 107 127 3 (POTENTIAL).  
 FT TRANSMEM 145 165 4 (POTENTIAL).  
 FT TRANSMEM 184 204 5 (POTENTIAL).  
 FT TRANSMEM 230 250 6 (POTENTIAL).  
 FT TRANSMEM 295 315 7 (POTENTIAL).  
 FT TRANSMEM 338 358 8 (POTENTIAL).  
 FT TRANSMEM 373 393 9 (POTENTIAL).  
 FT TRANSMEM 422 442 10 (POTENTIAL).  
 FT TRANSMEM 455 475 11 (POTENTIAL).  
 FT TRANSMEM 488 508 12 (POTENTIAL).  
 FT DOMAIN 509 525 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 525 AA; 54992 MW; 0183474MD2CC1CC6 CRC64;  
 Alignment Scores:  
 Pred. No.: 5,25e-08 Length: 525  
 Score: 327.00 Matches: 134  
 Percent Similarity: 39.58% Conservative: 92  
 Best Local Similarity: 23.47% Mismatches: 200  
 Query Match: 5.10% Indels: 145  
 DB: 1 Gaps: 21  
 US-09-759-143-110 (1-3410) x STP\_SPIOL (1-525)

QY 310 CTGCTGCGCACCGAAGCCAGCTCTTGCTGGTCAACCTGCTAACCCTTTGGCGCTGA 369  
 Db 25 ProThrThrProGlu-AlaGluAlaThrLeuLysLysLeu-----GlyLeuVa 41  
 QY 370 GGTGTGTTGGCCGAGGATCATACCTAT-----GTCCGCC 405  
 Db 41 lAlaSerValAlaAlaGlyValGlnPheGlyTrpAlaLeuGlnLeuSerLeuLeuThrPr 61  
 QY 406 TCTGCTGCTGGAGTGGGGTAGAGAGAGTTCATGACCATGCTGGCGATTTGCC 465  
 Db 61 oTyrValGlnLeuLeuGlyIleProHisThrTrpAlaAlaTyrIleTrpLeuCysGlyPr 81  
 QY 466 AGTGTGCTGGCTGGTCTGTCTCCGCTCTAGCTGACCCAGTGCACCTGGCGTGGAGC 525  
 Db 81 oileSerGlyMetIleValGlnProLeuValGlyTyrSerAspArgCysThrSerAr 101  
 QY 526 CTATGCGCCCGCCGCGCTTCATCTGGGCACTGCTTGGGCACTGCTGAGCCTCTT 585  
 Db 101 gPheGlyArgArgProPheIleAlaAlaGlyAlaAlaLeuAlaValAlaValG 121  
 QY 586 TCTCATCCCAAGGCGCGCTGCTAGCA-----GGGCTGCTGCTCCCGGATCC- 634  
 Db 121 yLeuIle-----GlyPheAlaAlaAspIleGlyAlaAlaSerGlyAspProThrGl 138  
 QY 635 -----AGGCCCCCTGGAGCTGGCACTGCTCATCTGGCGCTGGGCTGCTGGACTT 684  
 Db 138 yAsnValAlaLysProArgAlaIleAlaValPheValGlyPheTrpIleLeuAspVa 158  
 QY 685 CTGTGCGCAGGTGCTTCACTCCACTGGAGCCCTGCTCTGCACTTCCCG- --GA 741  
 Db 158 lAlaAsnAsnThrLeuGlnGlyProCysArgAlaLeuAlaLeuAlaAlaGlySe 178  
 QY 742 CCGGACCACTGTCGCCAGGCTACTCTGTATGCTTATGCTTATGCTTATGCTTATGCTG 801  
 Db 178 rClnThrLysThrArgTyrAlaAsnAlaPhePheSerPheMetAlaLeuGlyAsnIl 198  
 QY 802 CTGGGCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 837  
 Db 198 eGlyGlyTyrAlaAlaGlySerTyrSerArgLeuTyrThrValPheProPheThrLysTh 218  
 QY 838 CAGTCCCTGCGCCCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 895  
 Db 218 rAlaAlaCysAspValTyrCysAlaAsnLeuLysSerCysPhePheIleSerIleThrLe 238  
 QY 896 -CTCATCTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 954  
 Db 238 uLeuIleValLeuThrIleLeuAlaLeuSerValLysGluArgGlnIleThrIleAs 258  
 QY 955 CCCCACGAGCAGCAGAGGCTGCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1014  
 Db 258 pGluIleGlnGluGluAspLeuLysAsnArgAsnAsnSerSerGlyCys----- 275  
 QY 1015 CCGGCGCGCTTGGCTTTCGGAACCTGGCGGCTGCTTCCCGGCTGCTGCTGCTGCTGCTG 1074  
 Db 276 ---AlaArgLeuProPhe-----PheGlyClnLeuIleGlyAlaLeuLysAspLeu-- 291  
 QY 1075 CTGCGCGATGCGCCGACCTGCGCCGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1134  
 Db 292 -----ProLysProMetLeuIleLeuLeuLeuValThrAlaLeuAsnTrpIleAl 308  
 QY 1135 ACTCATGACCTTACGCTGTTTACAGGATTTCTGGGGGAGGGGCTGACAGGCGCT 1194  
 Db 308 aTrpPheProPheLeuLeuPheAspThrAspTrpMetGlyLysGluValTyrGlyGlyTh 328  
 QY 1195 GCCAGAGCTGAGCGGCGACCCGAGCGCGGACACTATGATGAAGCGCTGCGATGGG 1254  
 Db 328 r-----ValGlyGluGlyLysLeuTyrAspGlnGlyValHisAlaGl 342  
 QY 1255 CAGCTGGGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1314  
 Db 342 yAlaLeuGlyLeuMetIleAsnSerValValLeuGlyValMetSerLeuSerIleGluGl 362

QY 1315 GCTGGTGACGAGATTGCGC-----ACTCGAGCAGTCTATTGTCGCCAGTGTGGCAGCTTT 1368  
 Db 362 yLeuAlaArgMetValGlyGlyAlaLysArgLeuTrpGlyIleValAlaAsnIleLeu-- 381  
 QY 1369 CCCTGTGGCTCCCGGTGCCACATGCTTCCACAGTGTGCCCGTGGTGTGACAGCTTCAGC 1428  
 Db 382 -----AlaValCysLeuAlaMetThrVal---LeuValThrLysSerAl 395  
 QY 1429 CGCCCTCACCGGGTTTCACCTTCTCAGCCCTGCAGATCCTGCGCTFACAACACTGGCCTCCCT 1488  
 Db 395 aglu----- 396  
 QY 1489 CTACACCGGGAGAGAGGAGTGTTCCTGCGCCAAATACCGAGGGGACACTGGAGGTGTAG 1548  
 Db 397 -HisPheArgAspSerHisIleMet----- 405  
 QY 1549 CAGTGAGGACAGCCTGTATGACCATCTTCTCCAGGCGCTTAAGCCTGGAGCTCCTCTCCC 1608  
 Db 405 ----- 405  
 QY 1609 TAATGGACACGTGGTGTCTGGAGCAGTGGCTCTCCACCTCCACCCGCTCTGCGG 1668  
 Db 406 -----GlySerAlaValProProProProAla-----Gl 416  
 QY 1669 GGCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1728  
 Db 416 yValLysGlyGlyAlaLeuAlaIlePheAlaValLeuGlyIleProLeuAlaIleThrPh 436  
 QY 1729 G--GTTCGG-----GGCGGGGCGATCTGCTCT 1752  
 Db 436 eSerIleProPheAlaLeuAlaSerIlePheSerAlaSerSerGlySerGlyGlnGlyLe 456  
 QY 1753 GGACCTCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1795  
 Db 456 uSerLeuGlyValLeuAsnLeuAlaIleValProGlnMetPheValSerValThrSe 476  
 QY 1796 -----GCCCATCTCTGTTATGCGGCTCCAT 1821  
 Db 476 rGlyProTrpAspAlaMetPheGlyGlyGlyAlaValProAlaPheValValGlyAlaVa 496  
 QY 1822 TGTCAGCTGCGCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1852  
 Db 496 lAlaAlaThrAlaSerAlaValLeuSerPhe 506  
 RESULT 16  
 CA12\_MOUSE  
 ID CA12\_MOUSE STANDARD; PRT; 1459 AA.  
 AC P28481;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].  
 GN COL2A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91358489; PubMed=1885613;  
 RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon  
 structure, and alternative splicing.";  
 RL J. Biol. Chem. 266:16862-16869(1991).  
 RN [2]  
 RP SEQUENCE OF 1455-1459 FROM N.A.  
 RX MEDLINE=91274355; PubMed=2054384;  
 RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Specific hybridization probes for mouse type I, II, III and IX  
 collagen mRNAs";  
 RL Biochim. Biophys. Acta 1089:241-243(1991).  
 CC -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.  
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.





```

QY 1350 AAATAGACTGCTCGA-----GTGCGGAATCGCTGCACCGCGGTCCATCAGAGAG 1297
D 1351 |||||
D 462 LysArgGlyAlaArgGlyGluProGlyGlyAlaGlyProGlyGlyProGlyGlyAla 481
QY 1296 AAGA-----CCAGGAGATGCGCGACTGCAGAACAGCCCGCAGG 1258
D 1297 |||||
D 481 gGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGly---ProLysG 500
QY 1257 CTGCCCCATCCGAAGCCCTTCATCATAGTGTCTCCGGGCTCCG----- 1215
D 1258 |||||
D 500 yAlaProGlyGluArgGlyProSerGlyLeuAlaGlyProLysGlyAlaAsnGlyAspPr 520
QY 1214 ----TGCCGGCTCG-----CTCTGGCGACCCCTCGTATC 1183
D 1215 |||||
D 520 oGlyArgProGlyGluProGlyLeuProGlyAlaArgGlyLeuThrGlyArgProGlyAs 540
QY 1182 AGCC-----CCTCG 1174
D 1183 |||||
D 540 pAlaGlyProGlnGlyLysValGlyProSerGlyAlaProGlyGluAspGlyArgProG 560
QY 1173 CCCACGAAATCCGTGTAACAGCGTGAAGGTGATGATG-----CCATC 1129
D 1174 |||||
D 560 yProGlyProGlnGlyAlaArgGlyGlnProGlyValMetGlyPheProGlyProLy 580
QY 1128 CAGCTGCACAGCTCAGCCAGACGCGCG-----GCAGGGTGCAGGCGCATG 1081
D 1129 |||||
D 580 sGlyAlaAspGlyGluProGlyLysAlaGlyGluLysGlyLeuAlaGlyAlaProGlyLe 600
QY 1080 CGGCAGCACAGCTGTCACCGCGGAGCAGGCGCGCCAGGTTCGCGAAGCCAGCGG 1021
D 1081 |||||
D 600 uArgGlyLeuProGlyLysAspGlyGluThrGlyAlaAlaGlyProGlyProSerG 620
QY 1020 GCCCGCATGAGCAGCAGTGGCGGCGACAGGAGGCGCGACAGCCCTCTGCTGCTCG 961
D 1021 |||||
D 620 yProAlaGlyGluArgGlyGluGlnGlyAlaProGlyProSerGlyPhe----- 636
QY 960 GTGGGGCCAGCGCTGCTCTCAGCCACCA----- 930
D 961 |||||
D 637 -GlnGlyLeuProGlyProGlyProGlyProGlyGluGlyGlyLysGlnGlyAspGlnG 656
QY 929 -----GCAGTGTGGTGTACCGCAGGTGAGGAGATGAGGGTGAGC 889
D 930 |||||
D 656 yIleProGlyGluAlaGlyAlaProGlyLeuValGlyProArgGlyGluArgGlyPhePr 676
QY 888 AGGCCAAGAGGCACTCTCT-----CCTGCGTG 862
D 889 |||||
D 676 oGlyGluArgGlySerProGlyAlaGlnGlyLeuGlnGlyProArgGlyLeuProGlyTh 696
QY 861 CCCAGGT-----AGGGG-----CCAGGCGACGTGGTG 835
D 862 |||||
D 696 rProGlyThrAspGlyProLysGlyAlaAlaGlyProAspGlyProProGlyAlaGlnG 716
QY 834 TCCAGGTCATGCAGGCA-----GGAGGTAGCCCC 805
D 835 |||||
D 716 yProProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyIleAlaGlyProLy 736
QY 804 AGGCAGCCCCCAAGACTGTATCATGAAGCATAGACAGTAGTGGCTGGCGACAGTGGT-- 747
D 805 |||||
D 736 sGlyAspArgGlyAspValGlyGlyLysGlyProGlyGlyAlaProGlyLysAspGlyG 756
QY 746 -----CCGGGTCCC-----GGAAGAGGTCA 727
D 747 |||||
D 756 yArgGlyLeuThrGlyProIleGlyProGlyProAlaGlyAlaAsnGlyGluLysG 776
QY 726 GAGAGCAGGCGCTCCA---GTGGAGTGAAGCACA-----CC 694
D 727 |||||
D 776 yGluAlaGlyProProGlyProSerGlySerThrGlyAlaArgGlyAlaProGlyGluPr 796
QY 693 TGGCCACAGAAAGTCCAGCAGCAGCCCGC-----CCAGGATGACAGTGGC 649
D 694 |||||
D 796 oGlyGluThrGlyProProGlyProAlaGlyPheAlaGlyProProGly---AlaAspG 815

```

```

648 AGCTCCAGGGGCC---TGGATCCGGCGACAGAGCCCTGTAGCCAGCGCGCCCTTGGG 592
815 yGlnProGlyAlaLysGlyAspGlnGlyGluAlaGly----- 827
591 ATCAGAAAGAGGCTCAGCAGGATGCCCAAG----- 561
828 ---GlnLysGlyAspAlaGlyAlaProGlyProGlnGlyProSerGlyAlaProGlyPr 846
560 -----ACAGTGCCAGATCAAGGGCGCGCGCGCCATAGGCTCCA 520
846 oGlnGlyProThrGlyValThrGlyProLysGlyAlaArgGlyAlaGlnGlyProProGl 866
519 CGCCAGTGTCTACTGCTGAGCTAGGACGGGACACAGACAGGCG---CCAGCACTGA 463
866 yAlaThrGlyPhe-----ProGlyAlaAlaGlyArgValGlyProGlyAlaAs 883
462 CCAATCCCGCAGCAGCATGTCATCTCTCTACCCCTTCCAGCAGCAGAGGC 403
883 nGlyAsnProGlyProAlaGly-----ProProGlyProAlaGlyLysAs 898
402 GGCACATAGGTGATGCTCGCGCCAAACACACTCCAGCCAAAGTTAGCAGGTTGACC 343
898 pGlyProLysGlyValArgGlyAspSerGlyProProGlyArg-----AlaGlyAspPr 916
342 AGCAAGAGCTGGGCTTCCGGTGCC---GCACAGCGCGGCTCACCACAGCCCTCTGGACC 286
916 oGlyLeuGluGlyProAlaGlyAlaProGlyGlyLysGlyGluProGlyAspGlyPr 936
285 ATATGGG-----CCAGCGGGTAGGCTCAGGGGGCGGTTCCAGGCACACT--- 242
936 o-SerGlyLeuAspGlyProProGlyProGlnGlyLeuAlaGlyGlnArgGlyIleValG 956
241 -----CCAGAACTGCTCTGCTCTCGGCTCTGCTC 214
956 lLeuProGlyGlnArgGlyGluArgGlyPheProGlyLeuProGlyProSerGlyGluP 976
213 CAGAAAGTCCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 154
976 rGlyLysGlnGlyAlaProGlyAlaSerGlyAsp-----ArgGlyProProGly--- 992
153 ATTTCTGCCAGCCCTTGTGTCGCGTCCAGCTTCTCAGCCCATGCTCAACACCTCTGCT 94
993 -----ProValGly-ProProGlyLeuThrGlyProAlaGlyGluProGlyArg 1008
93 GTGGGCGACCTCAGTGGGACACGCTCATCTACTAGTCTCTGCGCGA----- 46
1009 GluGlySerProGlyAlaAsp-----GlyProProGlyArgAspGlyAlaAla 1024
45 -----GGCGCGGCTGTCCACCGGAGCC 22
1025 GlyValLysGlyAspArgGlyGluThrGlyAlaLeuGlyAlaProGlyAla 1041

```

```

RESULT 17
CA21_RANCA
ID CA21_RANCA STANDARD; PRT; 1355 AA.
AC O42350;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;
ON NCBI_TaxID=8400;
RX MEDLINE=97417499; PubMed=9272872;
RX Asahina K., Oofusa K., Obara M., Yoshizato K.;
RT "Cloning and characterization of the full length cDNA encoding alpha2
RT type I collagen of bullfrog Rana catesbeiana.";
RL Gene 194:283-289(1997).

```

-1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN (FIBRILLAR FORMING COLLAGEN).  
 -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM HYDROXYAPATITE.  
 -1- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D88764; BAA22380.1; -.  
 InterPro; IPR000087; Collagen.  
 InterPro; IPR000885; Fib\_collagen\_C.  
 Pfam; PF01391; Collagen; 18.  
 Pfam; PF01410; COL1; 1.  
 ProDom; PD000007; Collagen; 3.  
 ProDom; PD002078; Fib\_collagen\_C; 1.  
 SMART; SM00038; COLFI; 1.  
 Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 Glycoprotein; Collagen; Signal.  
 SIGNAL 1 22 POTENTIAL.  
 PROPEP 23 82 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).  
 CHAIN 83 1093 COLLAGEN ALPHA 2(I) CHAIN.  
 PROPEP 1094 1355 CARBOXYL-TERMINAL PROPEPTIDE (BY SIMILARITY).  
 CARBOHYD 1206 1206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 1256 1256 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SEQUENCE 1355 AA; 127643 MW; CB793AD5D6F4ID2A CRC64;

## Alignment Scores:

Pred. No.: 6.14e-08 Length: 1355  
 Score: 324.50 Matches: 304  
 Percent Similarity: 32.93% Conservative: 76  
 Best Local Similarity: 26.34% Mismatches: 435  
 Query Match: 5.21% Indels: 342  
 DB: 1 Gaps: 68

us-09-759-143-110 (1-3410) x CA21\_RANCA (1-1355)

QY 3029 GGGAAAGTTGGGGTAGGGAAGCTGGGGTAGGGGAAATTTTGGCAGTGCCTTCATC 2970  
 DB 226 GlySerAspGlySerGlyProValGlyProAlaGlyProIleGlySerAlaGly--- 244  
 QY 2969 AGCCAGTCTAGAGAGAGTAGAGGGAGT-----GGAGTGGGGGAACCCAG 2922  
 DB 245 AlaProGlyLeuProGlyAlaProGlyAlaLysGlyGluLeuGlyProAlaGlyAsnAsn 264  
 QY 2921 GCTGGCCCAAGAGAGGGGTGTAGGGAAGCCG-----TTGAGACCT 2877  
 DB 265 GlyProThrGlyAlaAlaGlyGlyArgGlyGluProGlyProGlySerLeuGlyPro 284  
 QY 2876 GAAGCCCAACCTCTACCTCTCCACACCCCTACCTTGGGTAAACAGCATTTTGGAAATTA 2817  
 DB 285 AlaGlyProGly-----AsnProGlyThrAsnGlyValAsnGly 298  
 QY 2816 TCATTTGGGATGAGTAGAATTTCCAAAGTCTCTGGGTGA-----GGCATTTGGGGGGCCAG 2761  
 DB 299 AlaLysGlyThrAlaGlyLeuProGlyValGlyGlyAlaProGlyLeuProGlyGlyArg 318  
 QY 2760 ---ACCCAGGAGAGAGATTCTGGCAATGATCATGCCCAATGA----- 2720  
 DB 319 GlyIleProGly-----ProIleGlyProAlaGlyProSerGlyAlaArgGlyLeuAla 336  
 QY 2719 -----CCAGGTATCTCAGGGAGCTGATTGTGGGATCCCCCAACCTACCCAAATATT 2666  
 DB 337 GlyAspProGlyIleAlaGlyGly-----LysGlyAsp----- 347

QY 2665 AGACACCAACACAGAAAAGCTAGCAATGGATTCCCTTCTACTTTTGTAAATAAAGTT 2606  
 DB 348 -----ThrGlySerLysGlyGluProGlySer----- 356  
 QY 2605 AAATATTAAATGCTGTCTGTGATGGCAACAGAGGACCAACAGGCCACATCCTG 2546  
 DB 357 -----ValGlyGlnGlnGlyProAlaGlyProSerGly 367  
 QY 2545 ATAAAGGTAAAGGGGGTGGATCAGCAAAAAGACAGTGTCTGTGGCTGAGGGGACCTG 2486  
 DB 368 GluGluGlyLysArgGlyProAsnGlyGluAlaGlySerSer----- 381  
 QY 2485 GTTCTGTGTGTGTCCTCAGGACTCTCCCTACAAATAGTCATATGTTCAAATCCC 2426  
 DB 382 -----GlyProSerGly----- 385  
 QY 2425 ATGGAGGAGTGTTCATCTCAGAAACTCCCATGCAAGAGCTACATTAAACGAAGTGCAG 2366  
 DB 385 ----- 385  
 QY 2365 GTTAAGGGCTTAGAGATGGGAAACAGCTGACTGAGTTTATTACGCTCCCAAAAACCT 2306  
 DB 386 AsnAlaGlyIleArg---GlyValProGlyThrArgGlyLeuProGlyProAspGlyArg 404  
 QY 2305 TCTTAGGTGTCTCAACTAGGAGGTAGCTGTAAACCTCAGCTGGGTAAATCCACT 2245  
 DB 405 AlaGlyGlyIleGlyProAlaGlySer-----ArgGlySerSerGlyProGly 421  
 QY 2245 CGAGAGTCCCGCATTCACAGTGCATGGAGCCCTTCT----- 2210  
 DB 422 AlaArgGlyProAsnGlyAspAlaGlyArgProGlyGluProGlyLeuLeuGlyAlaArg 441  
 QY 2209 GGCCTCCCTGTATTAAGTCCAGACTGAAACCCCTTTGGAAGGCTCCAGTCCAGGACCT 2150  
 DB 442 GlyLeuPro-----GlyPhe-----SerGlySerAsn 450  
 QY 2149 AGAGACTGGGGAGAGAGAGAGGGAGGGAGGGAGGGAGGGAGGGAGGGAGGGAGGGAGGG 2090  
 DB 451 GlyProGlnGly---LysGluGlyProAlaGlyProGlnGlyIleGluGlyArgSerGly 469  
 QY 2089 GCAGCA-----CAGGGTGGCAGCAGAGAGCCACATTTACTTTGGCAGCAACAACTGG 2036  
 DB 470 AlaAlaGlyProAlaGlyAlaArgGlyGluProGlyAlaIleGlyPheProGlyProLys 489  
 QY 2035 CGGCCAGCCCGCAGCCCTACAGAGCGGGAGCTGGGACCCAGTCCAGTCCAGGCA 1976  
 DB 490 GlyProAsnGlyGluProGlyLysAsnGlyAspLysGlyAsnGlnGlyProSerGlyAsn 509  
 QY 1975 GGCCTCCACCCCAATGTCTGGAAGCTTTTCTACGCTGAGTATTTGGCCCAAGTCGCTCT 1916  
 DB 510 ArgGlyAlaProGlyProAspGlyAsn----- 518  
 QY 1915 GTCAAACTACTCTGTGTAGCAAAAGTAAATGGCAGCAGACCCAGCCCTCGCCGACAGAC 1856  
 DB 519 -----AsnGlyAlaGln-----GlyProAla 525  
 QY 1855 CATATAGGAGTACAGACTGCTGAGCTGGACATGGAGCCCATAAACAGGATG---- 1800  
 DB 526 GlyLeuGlyGly---AlaThrGlyGluGlyGlnGlyPro---SerGlyAlaProG 544  
 QY 1799 -----GGGCCACTGGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1766  
 DB 544 LysPheGlnGlyLeuProGlyProGlyProGlyGluValGlyLys-----ProG 562  
 QY 1765 GATGGGAGGTCCAGCAGATGCCCGGCCG---GAACACCCCTGGCTCGGTGGGCTC 1709  
 DB 562 LysGluArgGly-----AlaProGlyAspPheGlyProPro---GlySerAlaGly-- 577  
 QY 1708 ACCCACCACACAGTACGAGGAGATCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1649  
 DB 578 -----ThrArgGlyGluArgGly-----AlaProGlyGluSerGlyAl 591

```

QY 1648 TGGAGCAGGCCACTGCTCCAGCAGCCACGCTGCCATTAGGAAGGAGCTCCAGGCTT 1589
Db 591 agly-----ProHisGlyProSerGlySerArgGlyProSerGly 604
QY 1588 AGGCCCTGGCAGGAAGCTGTCATCAGGCTGTCTCTACTGCTACGACCTCCAGTGTCCC 1529
Db 604 yAlaProGly-ProAspGlyGlnLysGlyGluProGlyAlaAlaGlyLeuAsnGlyGlyL 624
QY 1528 TCGGTATTGGGACAGAACCTCTCTCCCGTGTGTAGAGGAGGAGGAGGAGGAGGAGG 1469
Db 624 euGlyPro---SerGlyProAlaGlyLeuProGlyGluArgGly-----Thra 639
QY 1468 CAGGATCGAGGCTGAGAGGTGAACCGGTGAGGCGGCTGAAGCTGTCAACACAGGC 1409
Db 639 laGlyThrProGlyThrLys-GlyGluLysGlyAspAlaGlyAsnSerGlyAspTyrGly 658
QY 1408 CAC-----ACTGTGGACAGGCA 1391
Db 659 AsnProGlyArgAspGlyAlaArgGlyProAlaGlyAlaAlaGlyAlaProGlyProAla 678
QY 1390 TGTGGCACCGCAGCCAGGAAAGCTG-CCACACTGGGCCAAATAGACTGCTCGAGTGC 1332
Db 679 GlyGlyProGlyAspArgGlyGluSerGlyProAlaGlyProSer-----693
QY 1331 CGAATCGCTGCACAGCCGCTCCATGACAGAGAGAACAGGAGATGGCGCACTGCA 1272
Db 694 GlyValAlaGlyProArgGly-----AlaProGlyGluArgGlyGluAla 708
QY 1271 GGAACAGCC-----CCAGGCTGCCATCCGAGCCGCTTCATCATAGTGTCT 1227
Db 709 GlyProAlaGlyProThrGlyPheAlaGlyProGlyAlaAlaGlyHisThrGlyAla 728
QY 1226 TCCGGGCTCGG-----TGCCCGGCTCAGCTCTG 1197
Db 729 LysGlyAspArgGlyAlaLysGlyProLysGlyGluAlaGlySerProGlyProLeuGly 748
QY 1196 GCAGCCCTGGTACAGCCCTCGC---CCACGAAATCCCTGT---AAACAGCGGTGAAG 1143
Db 749 AlaHisGlySerAlaGlyProAlaGlyProAsnGlyProAlaGlySerThrGlyAlaArg 768
QY 1142 TCATGAGTG---CCATCCAGCTGCACAGCTCACAGGAGAGAGCCGCGC---GCAGGTGC 1089
Db 769 GlyAspAlaGlyProSerGlyAlaThrGlyPheProGlyProAlaGlyArgAlaGlyAla 788
QY 1088 GGGGCATCGCCGACAGCAGCTGTGTCAGCGGGAAGAGGCGCCAGCTTCGCGAAG 1029
Db 789 ProGlyProProGlyAsnValGlyProSerGlyProThrGlyHisProGlyLysAspGly 808
QY 1028 CCAAGCGGCGCCGATGGACAGCAGTGGGCGGACAAAG-----AGG 987
Db 809 SerArgGlyProArgGlyAspSerGlyProValGlyArgProGlyGluGlnGlnHis 828
QY 986 GGGCCGACAGCCCTCTG-----CTGGCTGGTGGGCGCCAGCGCTGCT 942
Db 829 GlyProValGlyLeuAlaGlyAspLysGlyProSerGlyGluAlaGlyProAlaGlyPro 848
QY 941 CCT-----CAGCCACAGCAGTGTGCTCTACGCGAG---TGAGGAAGATGAGG 894
Db 849 ProGlyAlaAlaGlyProSerGlyValLeuGlyAlaArgGlyLeuGlyLeuProGly 868
QY 893 TGAGCAGGCCAAAGAGGCTCTCT-----CCTGGTGCCCGAGGTAGG---GGGCCAGG 843
Db 869 ThrArgGlyGluArgGlyLeuProGlyGlyProGlySerAsnGlyGluProGlyProSer 888
QY 842 CACTGTGTCTCCAGTCAATGGCAGGAGCA-----GGTAGCCAGCAGCAGC 798
Db 889 GlyLeuAlaGlySerSerGlyProArgGlyProProGlySerValGlySerProGlyPro 908
QY 797 CCCCAAGACTGATCATGAAGGATAGCAGAGTAGGCTGGGACAGAGTGGTGGGTCC 738
Db 909 ValGlyHisSerGlyGluAlaGlyArgAspGlyHisProGlyAsnAspGlyPro---Pro 927
QY 737 GGAAGAGGTAGAGCAGGGGCTCCAGTGGAGTGAAGCACACCTGGCCACAGAGTCCA 678
Db 928 GlyArgAspGlyLeuProGlyAlaLysGlyGluArgGlyTyrProGlyAsnThrGlyPro 947
QY 677 GCA-----GCCCCAGGCCCA---GGATGAGCAGTCCAGCTCCAGGGCGCTGG 633
Db 948 SerGlyLeuAlaGlyAlaProGlyProAlaGlySerAlaGlyProAlaGlyLysSerGly 967
QY 632 GATCCGGGACAGCAGCCCTGTAGCCAGCCGCGCTGGGA-----TGA 588
Db 968 AsnArgGlyGluGlyGly-----ProSerGlyProAlaGlyIleThrGlyProSerGly 985
QY 587 GAAAGAGCTGACAGAGTGCACAGGACAGTCCAGATGAGGCGCGCGCGCCAT 528
Db 986 ProArgGlyProAlaGly---ProGlnGlyVal-----ArgGlyAspLysGlyGlu 1001
QY 527 AGCTCCACGCCAGTGTCTAGCTGGCTGACCTAGGAGGCGGACACAGA-----480
Db 1002 AlaGlyGluArgGlyAlaArgGlyLeuAspGlyArgLysGlyHisAsnGlyLeuSerGly 1021
QY 479 ---CCAGGCCAGCAGTGCACCAATGCCAGCAGCAGCATGTCATGACTTCTCTATCCC 423
Db 1022 LeuProGlyProSer-----GlyThrProGlyGluThrGlyProSerGlySerValGly 1039
QY 422 CCACCTT---CCAGCAGCAGCAGCGGCACATAGGTGATGCTCGCGCCAAACACACCTCCA 366
Db 1040 ProValGlyProArgGlyProSerGly-----ProSerGlyProPro 1053
QY 365 GCCCAAAG---TTAGCAGGTTGACACGACAAGA-----336
Db 1054 GlyLysGluGlyArgSerGlyHisProGlyAlaMetGlyProValGlyProArgGlyPro 1073
QY 335 GCTGGCTTTCGGT---GCCGACAGCGGCTCACCCAGCAGCTCTGGACCATAGTGG 279
Db 1074 AlaGlyPheThrGlyProAlaGlyProGlyProGlyProGlyProGly-HisAlaGly 1093
QY 278 GCCGCGCGGTAGGCTCAGGGGCG-----CGTTCCAGGCACTCC 240
Db 1093 yProSerGlyGlyGlyTyrAspGlyGlyAspGlyGlyGlyTyrArgAlaAspGlnPr 1113
QY 239 AGAA-----CTGCT 231
Db 1113 oGluArgLysProLysAspTyrGluValAspAlaThrLeuLysSerLeuAsnGlnIle 1133
QY 230 TCGTCTCGGCTCTGCTCCAGAGCTGCGGCTCTCTCTCTGCT-----TGCCGCAACT 177
Db 1133 eGluValIleLeuThrProGlyLysSerArgLysAsnProAlaArgThrCysArgAspLe 1153
QY 176 GCCTAGGAATCAGCCA-----GGGCGCCATTTCTGCCAGCCCTTTGGTGGCGG 129
Db 1153 uArgLeuSerHisProGluTyrThrSerGlyPheTyrTrpIleAspPro-----1169
QY 128 TCCAGCTTCTCAGCCCATGCTCAACACCTGTGCTG-----GTGGGCACTCAGT 78
Db 1170 -----AsnGln-GlyCysThrSerAspAlaIleArgValPheCysAspPheSerG 1187
QY 77 GGGACACGCTCTCATCCTCAGATCCT 52
Db 1187 yGluThrCysIleHisAlaAsnPro 1195

```

RESULT 18  
 CA12\_HUMAN  
 ID CA12\_HUMAN STANDARD; PRT; 1418 AA.  
 AC P02458;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(I) chain precursor [Contains: Chondrocalcin].  
 GN COL2A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=90067946; PubMed=2587287;  
RA Su M.W., Lee B., Ramirez F., Machado M., Horton W.;  
RT "Nucleotide sequence of the full length cDNA encoding for human type  
RT II procollagen.";  
RL Nucleic Acids Res. 17:9473-9473(1989).  
RN [2]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=87031574; PubMed=3021582;  
RA Nunez A.M., Kohno K., Martin G.R., Yamada Y.;  
RT "Promoter region of the human pro-alpha 1(II)-collagen gene.";  
RL Gene 44:11-16(1986).  
RN [3]  
RP SEQUENCE OF 432-1145 FROM N.A.  
RX Ramirez F.;  
RA Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 963-1418 FROM N.A.  
RX MEDLINE=85190534; PubMed=3857598;  
RA Cheah K.S.E., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;  
RT "Identification and characterization of the human type II collagen  
RT gene (COL2A1).";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).  
RN [5]  
RP SEQUENCE OF 1120-1398 FROM N.A.  
RX MEDLINE=85306861; PubMed=3840017;  
RA Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,  
RT Vuorio E.;  
RT "Construction and identification of a cDNA clone for human type II  
RT procollagen mRNA";  
RL Biochem. J. 229:183-188(1985).  
RN [6]  
RP SEQUENCE OF 1106-1418 FROM N.A.  
RX MEDLINE=88067771; PubMed=2825137;  
RA Elima K., Vuorio T., Vuorio E.;  
RT "Determination of the single polyadenylation site of the human pro  
RT alpha 1(II) collagen gene";  
RL Nucleic Acids Res. 15:9499-9504(1987).  
RN [7]  
RP SEQUENCE OF 1227-1289 FROM N.A.  
RX MEDLINE=86104139; PubMed=3002437;  
RA Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;  
RT "Isolation and partial characterization of genomic clones coding for  
RT a human pro-alpha 1 (II) collagen chain and demonstration of  
RT restriction fragment length polymorphism at the 3' end of the gene.";  
RL Biochemistry 24:6343-6348(1985).  
RN [8]  
RP SEQUENCE OF 1176-1226 FROM N.A.  
RX MEDLINE=84118798; PubMed=6320112;  
RA Strom C.M., Upholt W.B.;  
RT "Isolation and characterization of genomic clones corresponding to  
RT the human type II procollagen gene.";  
RL Nucleic Acids Res. 12:1025-1038(1984).  
RN [9]  
RP SEQUENCE OF 35-167 FROM N.A.  
RX MEDLINE=89233138; PubMed=2714801;  
RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;  
RT "Organization of the exons coding for pro alpha 1(II) collagen N-  
RT propeptide confirms a distinct evolutionary history of this domain of  
RT the fibrillar collagen genes.";  
RL Genomics 4:438-441(1989).  
RN [10]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=91184577; PubMed=2010058;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in collagen genes: causes of rare and some common diseases  
RT in humans.";  
RL FASEB J. 5:2052-2060(1991).  
RN [11]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97255959; PubMed=9101290;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-

RT associated collagen (type IX), and network-forming collagen (type X),  
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL Hum. Mutat. 9:300-315(1997).  
RN [12]  
RP VARIANT SER-1074.  
RX MEDLINE=90036909; PubMed=2572591;  
RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,  
RA Hollister D.W.;  
RT "Glycine to serine substitution in the triple helical domain of pro-  
RT alpha 1 (II) collagen results in a lethal perinatal form of short-  
RT limbed dwarfism.";  
RL J. Biol. Chem. 264:18265-18267(1989).  
RN [13]  
RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.  
RX MEDLINE=89266907; PubMed=2543071;  
RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;  
RT "Identification of the molecular defect in a family with  
RT spondyloepiphyseal dysplasia.";  
RL Science 244:978-980(1989).  
RN [14]  
RP VARIANT OSTEOARTHRTIS CYS-650.  
RX MEDLINE=90370826; PubMed=1975693;  
RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;  
RT "Single base mutation in the type II procollagen gene (COL2A1) as a  
RT cause of primary osteoarthritis associated with a mild  
RT chondrodysplasia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).  
RN [15]  
RP VARIANT OI-IV VAL-717.  
RX MEDLINE=91291136; PubMed=2064612;  
RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;  
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to  
RT valine substitution in osteogenesis imperfecta type IV. Detection of  
RT the mutation and prenatal diagnosis by a chemical cleavage method.";  
RL Biochem. J. 276:765-770(1991).  
RN [16]  
RP VARIANT OSTEOARTHRTIS CYS-650.  
RX MEDLINE=91086471; PubMed=1985108;  
RA Eyre D.R., Weis M.A., Moskowitz R.W.;  
RT "Cartilage expression of a type II collagen mutation in an inherited  
RT form of osteoarthritis associated with a mild chondrodysplasia.";  
RL J. Clin. Invest. 87:357-361(1991).  
RN [17]  
RP VARIANT HYPOCHONDROGENESIS GLU-984.  
RX MEDLINE=93054548; PubMed=1429602;  
RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,  
RA Cohn D.H., Eyre D.R.;  
RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha  
RT 1(II) chain produces hypochondrogenesis.";  
RL J. Biol. Chem. 267:22522-22526(1992).  
RN [18]  
RP VARIANT HYPOCHONDROGENESIS SER-705.  
RX MEDLINE=92262484; PubMed=1374906;  
RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,  
RA Ramirez F., Vitale E., Lee B.;  
RT "Characterization of a type II collagen gene (COL2A1) mutation  
RT identified in cultured chondrocytes from human hypochondrogenesis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).  
RN [19]  
RP VARIANT WS-II ASP-198.  
RX MEDLINE=93304428; PubMed=8317498;  
RA Koerkoe J., Ritvanemi P., Haataja L., Kaeerlaeinen H.,  
RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;  
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate  
RT for glycine alpha 1-67 and that causes cataracts and retinal  
RT detachment: evidence for molecular heterogeneity in the Wagner  
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";  
RL Am. J. Hum. Genet. 53:55-61(1993).  
RN [20]  
RP VARIANT SEMD CYS-840.  
RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,  
RA Eyre D.R.;  
RT "A dominant mutation in the type II collagen gene (COL2A1) produces



Db	417	ygluGluGlyLysArgGlyAlaArgGlyGluProGlyGlyValGlyProIleGlyProPr	437	Db	728	AlaGlyAlaAsnGlyGluLysGlyGluValGlyProGlyProAlaGlySerAlaGly	747
QY	906	AGGAAGATGAGGTGAGCAGGCAAGAGGC	871	QY	42	CGCGGGTGTCCACCGGA	25
Db	437	oGlyGluArgGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGlyPr	457	Db	748	AlaArgGlyAlaProGly	753
QY	870	TCTTGGTGGCCAGGT	814	RESULT 19			
Db	457	oLysGlyAlaProGlyGluArgGlyProSerGlyLeuAlaGlyProLysGlyAlaAsnG1	477	CA25_HUMAN			
QY	813	A-----GGTAGCCAGGAGCCCAAGACTGATCATGAAGGCATAGACAGTAGAGGCC	760	ID	CA25_HUMAN	STANDARD;	PRT; 1496 AA.
Db	477	YAspProGlyArgProGlyGluProGlyLeuProGlyAlaArgGlyLeuThrGlyArgPr	497	AC	P05997;		
QY	759	TGGCGACAGTGGTCCGGGTCCCGAAGAGGTGACAGAGCAGGCGCTCCAGTGGAGTGAAG	700	DT	01-APR-1988 (Rel. 07, Created)		
Db	497	oGlyAlaGlyPro-----GlnGlyLysValGlyProSerGlyAlaProGlyGluAspG1	516	DT	01-JAN-1990 (Rel. 13, Last sequence update)		
QY	699	CACACTGGCCAGAGTCCAGCAGCCCGCCG	664	DT	15-JUN-2002 (Rel. 41, Last annotation update)		
Db	516	YArgProGlyProProGlyProGlnGlyAlaArgGlyGlnProGlyValMetGlyPhePr	536	DE	Collagen alpha 2(V) chain precursor.		
QY	663	AGGATGAGCAGTGCAGCTCCAGGGGCTGGGATCCGGGCACACAGCAGCCCTGCTAGCCAG	604	GN	COL5A2		
Db	536	oGly-----ProLysGlyAlaAsnGlyGluProGlyLysAlaGly	549	OS	Homo sapiens (Human)		
QY	603	CCGCGCCTTGGATGAGAAAGAGCTCAGCAGATGCCCA-----AGGACAGTGGCCAG	550	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Db	550	-----GluLysGlyLeuProGlyAlaProGlyLeuArgGlyLeuPro---	563	OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
QY	549	ATGAAGGCGCGCGGCCCATAGCTCCAGCGCAGTGGTGTCACTGGGTGAGCTAGGAGC	490	OX	NCBI_TaxID=9606;		
Db	564	-----GlyLysAspGlyGluThrGlyAl	571	RN	[1]	SEQUENCE OF 1-463 FROM N.A.	
QY	489	GGGACACAGACAGGCGCCAGCA-----CTGGACCAATGCCAGC	451	RX	MEDLINE-89123368; PubMed-2914927;		
Db	571	agluGlyProGlyProAlaGlyProAlaGlyProAlaGlyGluArgGlyGluGlnGlyAlaProG1	591	RA	Woodbury D., Benson-Chanda V., Ramirez F.;		
QY	450	ACCATGTGTCAGTCTCTCTACCCCTCCAGCAGTGGTGTCACTGGGTGAGCTAGGAGC	490	RA	"Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms to the structural criteria of a fibrillar procollagen molecule.";		
Db	591	yPro-----SerGlyPheGlnGlyLeuProGly	600	RT	"The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrillar-forming collagens.";		
QY	390	ATGCTGCGCGCAACACACCTCCAGGCGCAAGGTTAGCAGGTTGACCAAGAGCTGG	331	RL	Nucleic Acids Res. 15:181-198(1987).		
Db	601	-----ProProGlyProProGlyGlu-----GlyLysProGlyAspGlnG1	615	RN	SEQUENCE OF 1227-1496 FROM N.A.		
QY	330	GCTTTCGGTCCGCGCAGGCGC-----	309	RX	MEDLINE-85289337; PubMed-3029669;		
Db	615	yValProGlyGluAlaGlyAlaProGlyLeuValGlyProArgGlyGluArgGlyPhePr	635	RA	Weil D., Bernard M.P., Gargano S., Ramirez F.;		
QY	308	-----GGCTCACCCAGAGCTCTGGA-----CCATAGTG	280	RA	"Complete primary structure of the human alpha 2 type V procollagen COOH-terminal propeptide.";		
Db	635	oGlyGluArgGlySerProGlyAlaGlnGlyLeuGlnGlyProArgGlyLeuPro-GlyT	655	RA	J. Biol. Chem. 260:11216-11222(1985).		
QY	279	GGCCAGGGGTAGGGCTCAGGGGCGCTCCAGGCATCCAGCACTGCTCGTCCGGCT	220	RT	"Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on the long arm of human chromosome 2.";		
Db	655	hrProGlyThrAspGlyProLysGlyAlaSerGlyPro-----A	668	RL	Genomics 3:275-277(1988).		
QY	219	CTGCTCCAGAGCTCGCGCTCTCTCTGCTGCCGCCAACTGCTAGGAATCAGCAG	160	RN	SEQUENCE OF 208-227.		
Db	668	laGlyProGlyAlaGlnGlyProProGlyLeuGlnGlyMetProGlyGluArgGlyA	688	RC	TISSUE-Placenta;		
QY	159	GGCGCATTTCTGCCAGCGCTTGT-----CCCGGTCCAGCTT	121	RX	Mann K.;		
Db	688	laAlaGlyIleAlaGlyProLysGlyAspArgGlyAspValGlyGluLysGlyProGluG	708	RX	"Isolation of the alpha 3-chain of human type V collagen and characterization by partial sequencing.";		
QY	120	CTCAGGCC-----ATGCTCAACACCTGCTGCTGTG	91	RA	Biol. Chem. Hoppe-Seyler 373:69-75(1992).		
Db	708	lyAlaProGlyLysAspGlyGlyArgGlyLeuThrGlyProIleGly-ProGlyPro	727	RA	[6]	SEQUENCE OF 288-297 AND 606-617.	
QY	90	GGGACCTCAGTGGGACAGCTCTCATCTAGATCCTGCG-----CGAGGC	43	RT	TISSUE-Bone;		
				RT	MEDLINE-94237164; PubMed-8181482;		
				RL	Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champiaud M.F.,		
				RN	Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;		
				RX	"Diversity in the processing events at the N-terminus of type-V collagen.";		
				RL	Eur. J. Biochem. 221:987-995(1994).		
				RN	[7]		
				RX	DISEASE.		
				RA	PubMed-9425231;		
				RA	Michalickova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;		
				RT	"Mutations of the alpha2(V) chain of type V collagen impair matrix assembly and produce Ehlers-Danlos syndrome type I.";		
				RL	Hum. Mol. Genet. 7:249-255(1998).		
				RN	[8]		





QY 2288 ACTAGGAGCTAGCTGTTAAACCTTGAGCTGGGTATATCCACCTGACAGTCCCGCATTC 2229  
Db 435 -----GlySerProGlyThrSerGlyProGly 444  
QY 2228 CAGTGCATGAGCCCTTGGCTCCCTGTATAGTCCAGACTGAACCCCTTGAAGG 2169  
Db 444 ySerAla---GlyProGlySerProGly---ProGlnGlySerThrGlyProGlnGly 462  
QY 2168 CTCCAGTCCAGGAGCCCTAGAGACTGGGAGAGAGG-----AGAGGGAC 2124  
Db 462 y---AsnSerGlyLeuProGlyAspProGlyPheLysGlyGluAlaGlyProLysGlyGly 481  
QY 2123 GCCCCAGCCCGCTGCTGAGCTAGCAGCTCAGCAGCAGCAGGAGGAGGAGAGAGCC 2064  
Db 481 uProGlyProHisGlyLeuGln----- 488  
QY 2063 ACATTACTTTGGCAGCAACAGAACTGGCGGCAGCCGCGGAGCCCATGGGCTAACAG 2004  
Db 489 -----GlyProLleGlyProProGlyGluGluGly 499  
QY 2003 GAGCGGGAGCTGGGA---CCAGTGGAGGAGCCCTCCACCCCAATGTGCTGGAAGTTT 1947  
Db 499 sArgGlyProArgGlyAspProGlyThrLeuGlyProPro----- 512  
QY 1946 TCTACGCTGAGTATTGGCGAGTCCCTCTGTGCAAAATACCTACCTGTGTAGCAAAAGTAA 1887  
Db 513 -----GlyProVal----- 515  
QY 1886 TGCGCAGCAGCCAGCGCTGGGCGCAGACACCATATAGCAGTGCAGACAGCTG----- 1835  
Db 516 -GlyGluArgGlyAlaProGlyAsnArgGlyPheProGlySerAspGlyLeuProGlyPro 535  
QY 1834 -----GCTAGCTGGCAATGGAGCCCATAAACAGGAGTGGGCGCAGCTGGGAGCA 1782  
Db 535 oLysGlyAlaGlnGlyGluArgGly-ProValGly-SerSerGlyProLysGlySerGln 554  
QY 1781 GGAAGCAGCTATCCAGATGGGAGGTCCAGCAGATGCCCGGCGGAGCAACCCCTGG 1722  
Db 555 Gly-----AspProGlyArgProGlyGlyPro-GlyLeuProGly 567  
QY 1721 CTTGGTGGCTCACCACACACACACAGTACGAGACATCAGAGGAGCA---GGCGCGCG 1665  
Db 567 yAlaArgGlyLeuThrGlyAsnProGlyValGlnGlyProGlyLysLeuGlyProLe 587  
QY 1664 AGAGCGGGGTGAGGTGGAGAGGCGCCACTGCTCCAGCAGCCAGCTGCTCCATTAGG- 1606  
Db 587 u-GlyAlaProGlyGluAspGlyArgProGlyProGly-----SerIleGlyI 604  
QY 1605 --AAGGAGCTCCAGGC---TTAGG---CCTGGCAGGAGCTGGTTCATCAGCTGCTCT 1554  
Db 604 leLysGlyProGlyThrMetGlyLeuProGlyProLys-GlySerAsnGlyAspPro 623  
QY 1553 CACTGTAGCAGCTCCAGTGTCCCTCGGTATT-----TGGCAGGAGCAACCTGCTTCT 1500  
Db 624 -----GlyLysProGlyGluAlaGlyAsnProGlyVal 634  
QY 1499 CCGGTGGTAGAGGAGG-----CCAGTGTGTAGGAGCA 1464  
Db 635 ProGlyLysArgGlyAlaProGlyLysAspGlyLysValGlyProGlyProGlyProGly 654  
QY 1463 TCTGCAGGCTGAGAGGTGAACCCGCTGAGGCGGCTGAAGCTGTACCAGCGGCACAC 1404  
Db 655 ProProGlyLeuArg-GlyGluArgGlyGluGlnGly---ProProGlyProThrGlyPhe 673  
QY 1403 TGTGGCAGGAGTGTGGCAGCGGAGCCAGCAGGAGGAAAG----- 1365  
Db 674 GlnGly-----HisProGlyProProGlyProProGlyGlyGlyPro 689  
QY 1364 -----CTGCCAGCTGGCCAAATAGACTGCTCGAAGTGGCAATGCTGCACCA 1317  
Db 690 GlyAspGlnGlyValProGlyGlyProGlyAlaVal-----GlyPro 703  
QY 1316 GCCGGTCCATGACCAGAGAGA---AGACCAGGAGATGGCGACTGCAGGAACAGCCCA 1260

Db 704 LeuGlyProArgGlyGluArgGlyAsnProGlyGluArgGlyGluProGlyIleThr--- 722  
QY 1259 GGCTGCCATCCGAACGC-----CTTCATCATAGTGTCTCCGGGCTCGG----- 1215  
Db 723 GlyLeuProGlyGluLysGlyMetAlaGlyGlyHisGlyProAspGlyProLysGlySer 742  
QY 1214 -----TGCCCGGCTCAGCTCTGGGACGCGCCCTGGT 1185  
Db 743 ProGlyProSerGlyThrProGlyAspThrGlyProGlyLeuGlnGlyMetProGly 762  
QY 1184 ACA-----GCCCTCGCCAGCAAAATCCGTGTAAACAGCGTGAAGTCA 1140  
Db 763 GluArgGlyIleAlaGlyThrProGlyProLysGlyAsp----- 775  
QY 1139 TGAGTGCCATCCAGCTGCACAGCTCAGCCAGGAAGAGCGCGGCGAGGTGGGGCATGC 1080  
Db 776 -----ArgGlyGlyIleGlyGlyLysGlyAla 784  
QY 1079 GGCAGCAGCTGTGTGACCGGGGAAAGCAGGCGCCAGGTTCGGAAGCCAGCGG 1020  
Db 785 GluGlyThrAlaGlyAsnAspGlyAlaGlyGlyLeuProGly-----ProLeuGly 801  
QY 1019 CCCGCATGGACAGCAGTGGGCGCAGAGGAGGCGGCGGACAGCCCTTCTGCTGGCTCGG 960  
Db 802 Pro-----ProGlyProAlaGlyLeu----- 809  
QY 959 TGGGCCCGCAGCTGCTCTCAGCCAGCAGCTGTGTGCTACGAGGTGAGGAAGA 900  
Db 810 --GlyGluLysGlyGluProGlyProArg-----GlyLeu 820  
QY 899 TGAGGTGAGCAGGCCAAAGAGGACACTCTCT----- 870  
Db 821 ValGlyProProGlySerArgGlyAsnProGlySerArgGlyGluAsnGlyProThrGly 840  
QY 869 -----CCTGGGTGCGCAGGTAGGGGG 849  
Db 841 AlaValGlyPheAlaGlyProGlnGlySerAspGlyGlnProGlyValLysGlyGlu--- 859  
QY 848 CCAGGCGACTGTGTCCAGTCAATGGCAGGAGGAGTCCAGCAGGAGCCAGCCCAAGAC 789  
Db 860 ProGlyGlu-----ProGlyGlnLysGlyAspAlaGlySerProGly---ProGlnGly 876  
QY 788 TGATCATGAGGCATAGACAGAGTAGGCTGGCGCAGCTGGTCCGGT----- 741  
Db 877 -----LeuAlaGlySerProGlyProHisGlyProAsnGlyValProGly 891  
QY 740 -----CCGGAAGAGGT 729  
Db 892 LeuLysGlyGlyArgGlyThrGlnGlyProProGlyAlaThrGlyPheProGlySerAla 911  
QY 728 CAGAGCAGGCGCTCCAGTGGAGTGAAGCAGACCTGGCCACAGAAAGTCCAGCAGCCCA 669  
Db 912 GlyArgValGlyProProGlyProAlaGlyAlaProGly-----ProAlaGlyPro 928  
QY 668 -----GCCCGCAGTATGAGCTGCCAGCTCCAGGCGCTGGGATCCGGGCGCAGCAGCC 615  
Db 929 LeuGlyGluProGlyLysGluGlyProGlyProArgGlyAspProGlySerHisGly 948  
QY 614 CTGTAGCAGCCCGGCTTGGGATGAGAAAGAGGCTCAGCAGGATCCCAAGGAGCAGTG 555  
Db 949 ArgValGly-----ValArgGlyProAlaGlyProProGly----- 960  
QY 554 CCCAGTGAAGGCGCGCGCGCATAGCTGCCAGCCAGTGTGCTGCTGAGCCTA 495  
Db 960 ----- 960  
QY 494 GGAGGGGACACAGCAGCCCGCAGCTGGACCAATGCCAGCAGCAGCAGTGTGTAAGCT 435  
Db 961 GlyProGlyAspLysGlyAspProGlyGluAspGlyGlnProGlyProAspGly----- 978  
QY 434 TCTCTCTACCCCTTCCAGCAGCAGGCGCGCATAGTGTGCTGCTGCGGCCAAC 375

Db 979 -----ProGlyProAlaGlyThrThrGlyGlnArgGlyIleVal----- 992  
 QY 374 ACACCTCCAGGCCAAGG-----TTAGCAGGTGA 345  
 Db 993 GlyMetProGlyGlnArgGlyGluArgGlyMetProGlyLeuProGlyProAlaGlyThr 1012  
 QY 344 CCAGCAGAGCTGGCTTCCGGTCCGCGCAGCAGGC-----GGC 306  
 Db 1013 ProGlyLysValGlyProThrGlyAlaThrGlyAspLysGlyProGlyProValGly 1032  
 QY 305 TCACCCACAGCTCGACCATAGTGGG---CCAGCGGGTAGGGCTCAGGGGGCCGCTTC 249  
 Db 1033 ProGlyLysAsnGlyProValGlyGluProGlyProGluGlyProAlaGlyAsnAs 1052  
 QY 248 AGGCACCTCCAGAACTGCTGCTCGCTGCTGCTCCAGAAAGCTCGCGCTCTCTCTCTCTG 189  
 Db 1052 pGlyThrProGly-----ArgaspGlyAlaValGlyGluArgGly----- 1065  
 QY 188 CTGCGCGCACTGCTAGGAATCAGCCAGCGCCCATTTCTGCCAGCCCTTTGTCGCGG 129  
 Db 1066 -----AspArgGlyAspProGly-----ProAlaGlyLeuPr 1076  
 QY 128 TCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGCGCACTCAGTGGGGACACGT 69  
 Db 1076 oGlySerGln-GlyAlaProGlyThrProGlyProValGly---AlaProGlyAspAlaG 1095  
 QY 68 CTCATCACTCAGATCCTGCG-----CGAGCGCGCG 39  
 Db 1095 LyGlnArgGlyAspProGlySerArgGlyProIleGlyHisLeuGlyArgAlaGlyLysA 1115  
 QY 38 GGCTGTACCCGCGCAGCGCCGCTGCAGG 10  
 Db 1115 rgGlyLeuProGlyProGlnGlyProArg 1124

RESULT 20  
 CA54 HUMAN  
 ID CA54\_HUMAN STANDARD; PRT; 1585 AA.  
 AC P29400; Q16126; Q16006;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 5(IV) chain precursor.  
 GN COL4A5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94165049; PubMed=8120014;  
 RA Zhou J., Leinonen A., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A5 gene.";  
 RL J. Biol. Chem. 269:6608-6614(1994).  
 RN [2]  
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.  
 RC TISSUE=Kidney;  
 RX MEDLINE=92316923; PubMed=1352287;  
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen  
 chain and identification of a single-base mutation in exon 23  
 converting glycine 521 in the collagenous domain to cysteine in an  
 Alport syndrome patient.";  
 RL J. Biol. Chem. 267:12475-12481(1992).  
 RN [3]  
 RP SEQUENCE OF 85-1685 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=90337990; PubMed=2380186;  
 RA Pihlajaniemi T., Pohjola-Erila E., Myers J.C.;  
 RT "Complete primary structure of the triple-helical region and the  
 carboxyl-terminal domain of a new type IV collagen chain, alpha  
 5(IV).";  
 RL J. Biol. Chem. 265:13758-13766(1990).  
 RN [4]

RP SEQUENCE OF 924-1685 FROM N.A.  
 RX MEDLINE=91169491; PubMed=2004755;  
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
 RT "Characterization of the 3' half of the human type IV collagen alpha  
 5 gene that is affected in the Alport syndrome.";  
 RL Genomics 9:1-9(1991).  
 RN [5]  
 RP SEQUENCE OF 914-1685 FROM N.A.  
 RX MEDLINE=90160375; PubMed=1689491;  
 RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,  
 Tryggvason K.;  
 RT "Identification of a distinct type IV collagen alpha chain with  
 restricted kidney distribution and assignment of its gene to the  
 locus of X chromosome-linked Alport syndrome.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
 RN [6]  
 RP SEQUENCE OF 1442-1471 FROM N.A.  
 RX MEDLINE=90252791; PubMed=2339699;  
 RA Myers J.C., Jones T.A., Pohjola-Erila E., Kadri A.S., Goddard A.D.,  
 Sheer D., Solomon E., Pihlajaniemi T.;  
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
 to the region of the X chromosome containing the Alport syndrome  
 locus.";  
 RL Am. J. Hum. Genet. 46:1024-1033(1990).  
 RN [7]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,  
 Marynen P.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).  
 RX MEDLINE=94133540; PubMed=8301933;  
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,  
 Cassiman J.-J., Marynen P.;  
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood  
 cells: a complex mutation in the COL4A5 gene of an Alport patient  
 deletes the NC1 domain.";  
 RL Kidney Int. 44:1316-1321(1993).  
 RN [9]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [10]  
 RP VARIANT AS SER-1564.  
 RX MEDLINE=91169492; PubMed=1672282;  
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,  
 Tryggvason K.;  
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a  
 conserved cysteine to serine in Alport syndrome.";  
 RL Genomics 9:10-18(1991).  
 RN [11]  
 RP VARIANT AS ARG-325.  
 RX MEDLINE=92303559; PubMed=1376965;  
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,  
 Tryggvason K., Gubler M.-C., Antignac C.;  
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5  
 (IV) chain associated with X-linked Alport syndrome: characterization  
 of the mutation by direct sequencing of PCR-amplified lymphoblast  
 cDNA fragments.";  
 RL Am. J. Hum. Genet. 51:135-142(1992).  
 RN [12]  
 RP VARIANT AS GLU-325.  
 RX MEDLINE=93244772; PubMed=1363780;  
 RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,  
 Rizzoni G.F., de Marchi M.;  
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
 glutamic acid in Alport syndrome.";  
 RL Hum. Mol. Genet. 1:127-129(1992).  
 RN [13]  
 RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
 RX MEDLINE=94010948; PubMed=8406498;

RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,  
 RA Tryggvason K., Hagsma-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
 van Oost B.A., Smeets H.J.M.;  
 RT "Identification of four novel mutations in the COL4A5 gene of  
 RT patients with Alport syndrome.";  
 RL Genomics 17:485-489(1993).  
 RN [14]  
 RP VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872  
 RP AND C-1241.  
 RA MEDLINE-95322976; PubMed-7599631;  
 RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
 RT "Detection of 12 novel mutations in the collagenous domain of the  
 RT COL4A5 gene in Alport syndrome patients.";  
 RL Hum. Mutat. 5:197-204(1995).  
 RN [15]  
 RP VARIANT AS ARG-1649.  
 RA MEDLINE-96213750; PubMed-8651292;  
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
 Denison J.C., Fain P.R., Gregory M.C.;  
 RT "A mutation causing Alport syndrome with tardive hearing loss is  
 RT common in the western United States";  
 RL Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [16]  
 RP VARIANTS AS.  
 RA MEDLINE-96213754; PubMed-8651296;  
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
 Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,  
 Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
 Savi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
 RT exons of the COL4A5 gene";  
 RL Am. J. Hum. Genet. 58:1192-1204(1996).  
 RN [17]  
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
 RP MET-1428  
 RA MEDLINE-97094179; PubMed-8940267;  
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,  
 Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broeyer M.,  
 Gubler M.-C., Antignac C.;  
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
 RT syndrome";  
 RL Am. J. Hum. Genet. 59:1221-1232(1996).  
 RN [18]  
 RP VARIANT AS ASP-1498.  
 RA MEDLINE-96233932; PubMed-8829632;  
 RA Tverskaya S., Bobrylina V., Tsalykova F., Ignatova M.,  
 Krasnopolskaya X., Evgrafov O.;  
 RT "Substitution of A1498D in noncollagen domain of a5(IV) collagen  
 RT chain associated with adult-onset X-linked Alport syndrome.";  
 RL Hum. Mutat. 7:149-150(1996).  
 RN [19]  
 RP VARIANT AS GLN-1677.  
 RA MEDLINE-97295089; PubMed-9150741;  
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
 RT "Common ancestry of three Ashkenazi-American families with Alport  
 RT syndrome and COL4A5 R1677Q";  
 RL Hum. Genet. 99:681-684(1997).  
 RN [20]  
 RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.  
 RA MEDLINE-98112435; PubMed-9452056;  
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
 Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
 Trivelli A., Pinciaroli A.R., Ragaio M., Rizzoni G.F., de Marchi M.;  
 RT "Missense mutations in the COL4A5 gene in patients with X-linked  
 RT Alport syndrome";  
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
 RN [21]  
 RP VARIANTS AS V-420; 456-P-P-458 DEL; D-573; D-624; D-635; 802-G--P-807  
 RP DEL; R-869; C-941; S-1030; S-1066; D-1143; R-1196; E-1261; S-1357  
 RP AND R-1649.  
 RA MEDLINE-99063529; PubMed-9848783;  
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,  
 Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,

RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;  
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
 RT Alport syndrome using PCR and direct DNA sequencing.";  
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
 RN [22]  
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;  
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
 RX MEDLINE-20030197; PubMed-10561141;  
 RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,  
 Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;  
 RT "Detection of mutations in the COL4A5 gene in over 90% of male  
 RT patients with X-linked Alport's syndrome by RT-PCR and direct  
 RT sequencing.";  
 RL Am. J. Kidney Dis. 34:854-862(1999).  
 RN [23]  
 RP VARIANT AS ARG-822.  
 RX MEDLINE-20025011; PubMed-10563487;  
 Alignment Scores:  
 Pred. No.: 7e-08 Length: 1685  
 Score: 323.00 Matches: 284  
 Percent Similarity: 30.90% Conservative: 59  
 Best Local Similarity: 25.59% Mismatches: 339  
 Query Match: 5.03% Indels: 428  
 DB: 1 Gaps: 65  
 US-09-759-143-110 (1-3410) x CA54\_HUMAN (1-1685)  
 QY 2 GGAACAGCGCTGCACGCGTGGCTCGGGTGACAGCGCGCGCCT-----CGG 49  
 Db 486 GlyThrGlyIleSerGly---ProProGlyGlnProGlyLeuProGlyLeuProGlyPro 504  
 QY 50 CCAGGATCTGAGTGATGACAGCTGCCACCTGAGTGCCCGCCACAGCAGCAGCGTGTGAG 109  
 Db 505 ProGlySer-----LeuGlyPheProGlyGlnLysGly---Glu 516  
 QY 110 CATGGCTGAGAGCTGGACCGGCACCAAGAGGCTG-----GCAGAAATG 154  
 Db 517 LysGlyGlnAlaGlyAlaThrGlyProLysGlyLeuProGlyIleProGlyAla-ProG 536  
 QY 155 GCGCCCTGGCTGATCTCTAGGCGAGTGGCGGCGAGCAAGGA----- 194  
 Db 536 yAlaProGly---PheProGlySerLysGlyLeuProGlyAspIleLeuThrPheProG 555  
 QY 195 -----GGAGAGCGCGCAGCTTCTGGAGCAGAGCGCGAGAGCAAGAGTTC 238  
 Db 555 yMetLysGlyAspLysGlyGlnLeuGlySerProGlyAlaPro-----GlyLeuPr 572  
 QY 239 TGGAGTGCCCTGAACGGCCCTTACCCGCTGACCCCTGCGCCCTACTATGTTCCAGAGCGTG 298  
 Db 572 oGlyLeuProGlyThrProGlyGlnAspGlyLeuProGlyLeuProGlyProLys- 590  
 QY 299 TGGGTGAGCGG-----CCTGCTGGCGCAGCGGAAGCCAGCTC 337  
 Db 591 --GlyGluProGlyGlyIleThrPheLysGlyGluArgGlyProProGlyAsnProGlyL 610  
 QY 338 TTGCTGCTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTTGGCGCAGCATCACCTAT 397  
 Db 610 euProGlyLeuProGlyAsnIleGlyProMetGly----- 621  
 QY 398 GTGCCCGCTCTGCTGTGGAGTGGGGGTAGAGAGAGTTCATGACCATGGT----- 450  
 Db 622 -----ProProGlyPheGlyProProGlyProValGlyGlyLysGlyIleGlnG 638  
 QY 451 -----GCTGGGCATTTGCTCCAGTCTGGCTGTGTGCTCCCGCTCTCTAGCTCAGCC 505  
 Db 638 lyValAlaGly-----AsnProGlyGlnProGlyIle---ProGlyProLys----- 652  
 QY 506 AGTGACCACTGGCGTGGACGCTATGCGCGCGCGCGCTTTCATCTGGGCACCTGTCTCTG 565  
 Db 563 -----G 653  
 QY 566 GGCATCTGCTGAGCCTCTTTCTTCATCCCAAGGCGCGCTGGCTAGCAGGGCTGCTGTGC 625

[illegible]

2657	QY	TTGGTCTCTAATATTGGGTAGGGTGGGGATCCCAACAATCAGTCCCTGAGATAGC	2716
1164	Db	-----GlyGluLysGlyLysProGlyGlnAspGly-----	1173
2717	QY	TGGTCATTGGCTGATCATCCCA-----GAATCTCTTCTCTCTGGGGCTGG	2764
1174	Db	-----Ile-ProGlyProAlaGlyGlnLysGlyGluProGly-----	1185
2765	QY	CCCCCAAAATGCTTAACCCAGGACGCTTGGAAATCTACTATCATCCCAATGATAATCCA	2824
1186	Db	GlnProGlyPheGlyAsnProGlyPro-----	1195
2825	QY	AATGCTGTTACCCAAGTTAGGCTGTGAAGGAAGGTAGAGGCT-----GGGGCTTCAG	2878
1196	Db	-----GlyLeuProGly--LeuSerGlyGlnLysGlyAspGlyG	1208
2879	QY	GTCTCAAGGCTTCCCTAACCAACCCCTCTCTCTTGTGGCCAGGCTGGTCCCCCCTTC	2938
1208	Db	lyLeuProGlyIleProGlyAsnProGlyLeuProGlyPro-LysGlyGluProGlyPhe	1227
2939	QY	CAC-----TCCCCCTACTCTCTAGGACTGGGCTGATGAAGGCAC	2980
1228	Db	HisGlyPheProGlyValGlnGlyPro-----	1236
2981	QY	TGCCCAAAATTTCCCTACCCCAACTTTCCTACCCCAACTTTCCTACCCAGCTCCA	3040
1237	Db	-----ProGlyProProGlySerProGlyProAlaLeuGluGlyProLysGly	1252
3041	QY	CAACCC	3046
1253	Db	AsnPro	1254

## RESULT 21

KW	Glycoprotein; Collagen.
FT	MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID
FT	MOD_RES 9 9 CONVERTED TO AN ALDEHYDE GROUP THAT IS INVOLVED IN CROSS-LINKING.
FT	MOD_RES 103 103 HYDROXYLATION.
FT	CARBOHYD 103 103 O-LINKED (GAL. .).
FT	MOD_RES 115 115 HYDROXYLATION (POTENTIAL).
FT	MOD_RES 124 124 HYDROXYLATION (POTENTIAL).
FT	NON_CONS 145 146
FT	MOD_RES 274 274 HYDROXYLATION (POTENTIAL).
FT	MOD_RES 346 346 HYDROXYLATION (POTENTIAL).
FT	MOD_RES 424 424 HYDROXYLATION (POTENTIAL).
FT	MOD_RES 496 496 HYDROXYLATION (POTENTIAL).
FT	MOD_RES 658 658 HYDROXYLATION (POTENTIAL).
FT	MOD_RES 670 670 HYDROXYLATION (POTENTIAL).
FT	MOD_RES 726 726 ONLY HYDROXYLATED PRO IN POSITION X (IN THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
SQ	SEQUENCE 779 AA; 70346 MW; E554A7FF084283DL CRC64;
 Alignment Scores:	
Pred. No.:	9.3le-08 Length: 779
Score:	321.00 Matches: 229
Percent Similarity:	31.12% Conservative: 38
Best Local Similarity:	26.69% Mismatches: 269
Query Match:	5.16% Indels: 324
DB:	1 Gaps: 51
 US-09-759-143-110 (1-3410) x CALL_BOVIN (1-779)	
QY	2365 GTTAAGGGCTTAGA-----GATGGAAACCAAGTGACTGAGTTATTCA 2321 :::
Db	102 MetLysGlyHisArgGlyPheSerGlyLeuaspGlyAlaLysGlyasp-----Ala 118 
QY	2320 GCCTCCAAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCCTGAGC 2261 
Db	119 GlyProAlaGlyProllys----- 124
QY	2260 CTGGGTAATCCA-----CCTCGAGAGTCGCCGCATTCACGTGATGAGGCCCTTCTGGC 2207

Db 125 ---GlyGluProGlySerProGlyGluAsnGlyAlaProGlyGlnMetGlyProArgGly 143  
QY 2206 CTCCTGTATAGTCCAGACTGAACCCCTTGGAGGCTCCAGTCAGCCAGCCCTAGA 2147  
Db 144 LeuPro-----GlyPheProGlyProLysGlyAlaAlaGlyGluProGly 158  
QY 2146 GACTGGGAGAGAGAGGAGCCGCCAGCCCA---GCTGTGACGTACGCACCTCA 2090  
Db 159 LysAlaGlyGlu---ArgGlyValProGlyProGlyAlaValGlyProAlaGlyLys 177  
QY 2089 GCACACAGGGTGCACAGAGCCACATTTACTTTGGCAGCAACAACTGGCGGCA 2030  
Db 178 AspGlyGluAlaGlyAlaGlnGlyPro----- 186  
QY 2029 GCCCGGAGCCCATGGGCTAACAGAGCGGGGAGCTGGGA----- 1988  
Db 187 -----ProGlyProAlaGlyProAlaGlyGluArgGlyGluGlnGlyProAlaGly 203  
QY 1987 -----CCAGTGTAGGAGGCGCTCCACCCCAATGTGTGGAAGT 1949  
Db 204 SerProGlyPheGlnGlyLeuProGlyProAlaGlyProGlyGlu---AlaGlyLys 222  
QY 1948 TTTCTAGCTGAGTATTGGCCAAAGTCGCTCTTGTCAATACTACCTGTGTAGCAAGTA 1889  
Db 222 ----- 222  
QY 1888 AATGGGACACAGAC---CCAGGCTCGGCAGACACCATATAGGAGTGCACAGACTGC 1833  
Db 223 ProGlyGluGlnGlyValProGlyAspLeuGlyAlaPro----- 235  
QY 1832 TGAGCTGGACAATGGAGCCCATAAACAGGATGGGGCCACCTGGGACAGCAGGA---AGG 1776  
Db 236 -----GlyProSerGlyAlaArgGlyGluArg 244  
QY 1775 CACTATCCAGATGGGAGTCCAGCAGATGCCCCGGCCCGGACCAACCTCG---GCTCG 1717  
Db 245 GlyPheProGlyGluArgGlyValGluGlyProGlyProAlaGlyProArgGlyAla 264  
QY 1716 GTGGGCTACCCACACC---ACACGTAGGAGACATACAGGAGGCGCCCGGAG 1663  
Db 265 AsnGlyAlaProGlyAsnAspGlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGly 284  
QY 1662 AGCGCGGTGGAGTGGGAGCAGGCCATGCT-----CCA 1627  
Db 285 SerGlnGlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuPro 304  
QY 1626 GCACCCACGTGTCATTAGG-----AAGGAGCTCCAGGCTTAGGCTGCG 1579  
Db 305 GlyProLysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGly---AlaProGly 323  
QY 1578 AGGAAGCTGTATCAGGTGCTCCTAGTGTAGCAGCTCCAGTGTCCCTCGGTAAT-- 1521  
Db 324 LysAspGlyValArgGlyLeuThrGlyProIleGlyProPro-GlyProAlaGlyAlaPr 343  
QY 1520 -----TGGCAGGAACACCTGTTCTCCGGTGG 1492  
Db 343 OglyAspLysGlyGluAlaGlyProSerGlyProAlaGlyThrArgGlyAlaProGlyAs 363  
QY 1491 TAGAGGAGGCGAGTGTGTAGGCGAGGATCTCGAGGCTGAGAGGTGAACCCGCTGAGG 1432  
Db 363 pArgGlyGluPro-----GlyProProGly----- 371  
QY 1431 GCGGCTGAAGCTGTACACCGCCACACTGTGGGACAGGCATGTGGCACCGGACCCACA 1372  
Db 372 -----ProAl 373  
QY 1371 GGGAAAGTGCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACAGCGG 1312  
Db 373 agly-----PheAlaGlyProProG 380  
QY 1311 TCCATGACACAGAGAACAGCAGGAGATGGCGCACTGCAGGAACAGCCCGCTGCC 1252  
Db 125 --- 125

380 y-----AlaAspGlyGlnProGly-----AlaLysGlyLeuPr 391  
1251 ATCCGAACGCTTCATCATAGTCTCCGGGCTCGGTCCCGGCTCAGCTCTGGGCAGG 1192  
391 OglyAspAlaGlyAlaLysGlyAspAlaGlyProGlyProAlaGlyProAlaGlyPr 411  
1191 CCTGTGTACA-----GCCCTCGCCCAACCAATCCGTGTAAACACAGCGNG 1147  
411 OProGlyProIleGlyAsnValGlyAlaProGlyPro-----LysGlyAlaAr 427  
1146 AAGGTATCATGTCCTCAGCTGCACAGCTCAGCCAGAGAGCGGCGCAGGTGCGG 1087  
427 gGlySerAlaGlyProProGlyAlaThrGlyPhePro-----G 440  
1086 GGCATGCGCGCAGCACAGCT-----GGTGCACCGGGGAAGCAGGCGCCAGGTTCGG 1033  
440 yAlaAlaGlyArgValGlyProProGlyProSerGlyAsnAlaGlyProProGly- 458  
1032 AAAGCAAGCGCGCGGCGCATGAGCAGTGGGCGCAGAGAGGCGGCGCAGACGCT 973  
459 ---ProProGlyProAlaGlyLysGluGlySerLysGlyProArgGlyGluThrGly- 476  
972 TCTGTGCTGCTGCGTGGGCGCCAGCCTGCTCCTCAGCCAGCAGCAGTGTGCTCTAG 913  
477 ProAlaGlyArgProGlyGluValGlyProProGlyProPro----- 490  
912 CAGGTGAGGAAGATGAGGTGAGCGCAAGAGGACACTCTCT----- 870  
491 -----GlyProAlaGlyGluLysGlyAlaProGlyAlaAspGlyProAl 505  
869 -----CCTGGTCCCGAGT-----AGGGG----- 849  
505 aglyAlaProGlyThrProGlyProGlnGlyIleAlaGlyGlnArgGlyValGlyLe 525  
848 -CCAGGCGACTGGTCTCCAGTCAATGCAGCAGGA-----GCTAGCCCGCAGCGCC 796  
525 uProGly-----GlnArgGlyGluArgGlyPheProGlyLeuPr 538  
795 CAAGACTGATCATGAAGCATAGACAGTAGGCTCTGGCAGACTGCTCGGTCGCGG 736  
538 OglyPro-----SerGlyGluProGlyLysGlnGlyProSerGlyAl 552  
735 AAGAGTCAGAGACAGGCGCTCCAGTGGAGTGAACACACCTGCGCCACAGCACTCA-- 678  
552 aSerGlyGluArg--GlyProProGlyProMetGlyProProGlyLeuAlaGlyPro 571  
677 ---GCAGCCCGCAGCAGTGCAGTGCAGTCCAGCGGCGC---TGGGATCCGG 625  
571 OglyGluSerGlyArgGluGly-----AlaProGlyAlaGluGlySerProG 587  
624 CACAGCAGCCTGTAGCCAGCGCGCTTGGGATGAGAAGAGGCTCAGCAGGATGCC 565  
587 yArgAspGlySerProGlyAlaLysGlyAspArgGlyGluThrGlyProAlaGlyAlaPr 607  
564 AAGGACAGTGCAGATGAAGCGCGCGGCGCATAGCTCCACCGCAGTGTGCTACTG 505  
607 OglyProPro----- 610  
504 GCTGAGCTTAGGAGCGGACAGACAGCCAGCGCCAGCACTGGACCAATGCCAGCAGCATG 445  
611 -----GlyAlaProGlyAlaProGlyProVal-----GlyProAlaGlyLy 624  
444 GTCA-----TGAATCTCTCTACCCCACTT---CCAGCAGCAGAGCGGCG 400  
624 sSerGlyAspArgGlyGluThrGlyProAlaGlyProIleGlyProValGlyProAlaG 644  
399 ACATAGTGTATGCTCGCGCAACACACCTCCAGGCGCAAGG----- 357  
644 yAlaArgGlyProAlaGlyProGlnGlyProArgGly\*\*LysGly\*\*\*ThrGly\*\*\*\* 664  
356 -----TTAGCAGGTTCAGCAGCAAGAGTGGCTTTCGGT---GCCAGCAG 313  
664 \*Gly\*\*\*ArgGlyIleLysGlyHisArgGlyPheSerGlyLeuGlnGlyProProGlyPr 684





```
Db 470 euProGlyIleasp --- GlyArgProGlyProIleGlyP 482
QY 2235 CCATTCACAGTCAGGAGCCCTTCTGCGCTCCCTGTATAGTCCAGACTGAACCCCT 2176
Db 482 roAlaGlyAlaArgGlyGluProGlyAsnIleGlyPhe ---ProG 496
QY 2175 TGAAGGCTCCAGTCAGGACGCCCTAGAGACTGGGGAGAGAG --- 2132
Db 496 lyProLysGlyProThrGlyAspProGlyLysAsnGlyAspLysGlyHisAlaGlyLeuA 516
QY 2131 -----AGAGGAGCGCCAGCCAGCCAGCTGTCAGCTACGACCTCAGCAGCAGG 2080
Db 516 laGlyAlaArgGlyAlaProGlyProAspGly ---AsnAsnGlyAlaGlnG 532
QY 2079 GTGGCAGCAGAGACCATACTTTGTCAGCAACAGAAACTGGCGCCAGCCCGGAGC 2020
Db 532 ly-----ProProGlyP 536
QY 2019 CCATGGGCT---AACAGAGCGGGAGCTGGGACCCAGTGGAGCGCCCTCCA --- 1967
Db 536 roGlnGlyValGlnGlyLysGlyGlnGlnGlyPro-----AlaGlyProProGlyP 554
QY 1966 -----CCCCAATGCTGGAAGTTTCTACGCTAGTATTGGCAAGTCGCTC 1918
Db 554 heGlnGlyLeuProGlyProAlaGlyThr -----AlaGlyGluValGly- 568
QY 1917 TTGCAATACTACTGTGTAGCAAAAGTAATGTCGACAGCCAGCCAGCTGGCGGAGC 1858
Db 569 -----Lys---ProGlyGluArgGlyLe 575
QY 1857 ACCATATAGGAGTACAGACTGCTGAGCTGGACAAATGGAGCCCAATAACAGGATGGG 1798
Db 575 uProGly-----GluPheGlyLeuProGlyProAlaGlyProArgGlyGluArgG 592
QY 1797 GCCACCTGGGACAGCAGAGGCACTATCCA -----GGATGGCGAGTCCAGG 1750
Db 592 yProProGlyGluSerGlyAlaAlaGlyProSerGlyProIleGlySerArgGlyProse 612
QY 1749 CAGATGCCCGCCCGGAAC -----CACCTGGCTCGGTGGCTCAGCCACC 1702
Db 612 rGlyProProGlyPro-AspGlyAsnLysGlyGluProGlyVal-LeuGlyAlaProGly 631
QY 1701 ACCACAGCTAGGAGATCACAGGCAGAGCCCGCAGAGCCGGGTGAGTGGAGC 1642
Db 632 Thr-----AlaGlyAlaSerGlyPro 638
QY 1641 AGGCCACTGCTCCAGCAGCCAGCTGTTCATTAGGGAAGGAGCTCCAGGCTTAGGCGCT 1582
Db 639 GlyGlyLeuProGlyGlu-----ArgGlyAlaAlaGlyIle---Pro 651
QY 1581 GCGAGGAGCTGGTCATCAGGCTGCTCTACTGCTAGCACCTCCAGTGTCCCTCGGTAT 1522
Db 652 Gly-GlyLysGlyGlyLysGly -----GluThrGlyLeuArgG 664
QY 1521 TTGGCAGGAGCACTCTTCTCCCGTGTGTAGAGGAGCCAGTGTGTAGGCGCAGATC 1462
Db 664 yGluIleGlyAsnProGlyArgAspGlyAlaArgGlyAlaPro-----GlyAl 680
QY 1461 TGCAGGCTCAGAAGGTGAACCCGGTGAGGGCGCTGAAGCTGTCCACCGCCAGCAGCTG 1402
Db 680 aMetGly-----AlaProGlyProAlaG 688
QY 1401 TGGCAGCAGGATGTGGCAGCGGAGCCAGCCAGGAAAGCTGCCACACTGCCAAATAGACT 1342
Db 688 yAlaThrGly-----AspArgGlyGlu----- 695
QY 1341 GCTCGAGTGCCGAATCGCTGCACAGCCGCTCATCATCAGCAGAGAGACCGAGAGATG 1282
Db 696 -----AlaGlyProAlaGlyPro----- 701
QY 1281 GCGACTGTCAGGAACAGCC-----CCAGGCTGCCCATCCGAACGCTTTCATCATAGTGT 1228
Db 1017 -----ArgGly-----LeuProGlyLeuLys 1023
Db 702 -----AlaGlyProAlaGlyProArgGlyThrProGlyGluArgGlyGluValGlyP 719
QY 1227 CTCGGGCTCTCGG-----TGCCCGCTCAGCTCTGGGCACGCCCTGTGTAC----- 1182
Db 719 oAlaGlyProAsnGlyPheAlaGlyProAlaGlyAlaAlaGlyGlnProGlyAlaLysG 739
QY 1181 -----GCCCCTCGC-----CCAGAAATCCGTG 1159
Db 739 yGluArgGlyThrLysGlyProLysGlyGluAsnGlyProValGlyProThrGlyProI 759
QY 1158 TAAACACCTCTGAAGGTTCATGTCATCCAGCTCAGCTGCACAGCTCAGCCAGAGACCGG 1099
Db 759 eGlySerAla-----GlyProSerGlyProAsnGlyProProGlyProAlaG 775
QY 1098 CGCA-----GGTGGCGGCATGCGGCAGCAGCT-----GGTGCAGCCGGGAGC 1051
Db 775 ySerArgGlyAspGlyGlyProGlyAlaThrGlyPheProGlyAlaAlaGlyArgTh 795
QY 1050 AGGCGCCAGCT-----TCCGGAAGCCAGGCGCCGCGCATGGACAGCAG 1003
Db 795 rGlyProProGlyProSerGlyIleThrGlyProGlyProGlyProGlyAlaAlaGlyL 815
QY 1002 TGGGCGCACAAGG-----AGGGGCGCAGACAGCCCTTCTGCTGCTCGCTG 958
Db 815 sGluGlyLeuArgGlyProArgGlyAspGlnGlyPro-----ValGlyArgTh 831
QY 957 GGGCCAGCGCTCCTCAGCCAGCAGCTGTGGCTGTCTACGAGGTGAGGAGATG 898
Db 831 rGlyGluThrGlyAlaSerGlyProGlyPheThr-----GlyGlyL 846
QY 897 AGGTGAGCAGCCCAAGAGGCACT-----CCTCTGGTGCCCAAGT----- 855
Db 846 sGlyProSerGlyGluProGlyThrAlaGlyProGlyThrProGlyProGlnGlyLe 866
QY 855 ----- 855
Db 866 uLeuGlyAlaProGlyIleLeuGlyLeuProGlySerArgGlyGluArgGlyLeuProG 886
QY 854 -----AG 853
Db 886 yValAlaGlySerValGlyGluProGlyProLeuGlyIleAlaGlyProProGlyAlaAr 906
QY 852 GGGCCAGCGCTCTGTGTCCAGTCAATGGCAGCAGGAGTAGCCAGCCAGCCAGCCCA 793
Db 906 gGlyPro-----ProGlyAlaValGlyAlaProGlyValAsnGlyAlaProG 922
QY 792 AGACTGATCATGAAGCATAGACAGTAGGCTCGGCAGACAGTGTGTGCGGTCCGGAAG 733
Db 922 yGlu-----AlaGlyArgAspGlyAsnProGlyAsnAspGlyPro---ProGlyAr 938
QY 732 AGGTGAGCAGCAGCGCTCCAGTGGAGTGAAGCAGCAGCTGCCACAGACAGTCCAGCAGC 673
Db 938 gAspGlyGlnAlaGlyHisLysGlyGluArgGlyTyrProGlyAsnIleGlyProValG 958
QY 672 CCCAGCCCGAGTATGAGTCCAGCTCCAGGCGCTGGGATCCGGCAGCAGCAGCCCT 613
Db 958 y-----AlaValGlyAlaProGlyProHisGlyPro----- 968
QY 612 GCTGACCGCCCGCTTGGGATGAGAAGAGGCTCAGCAGATGCCCAAGACAGTGC 553
Db 969 -----ValGlyProThrGly---LysHisGlyAsnArgGlyGluPro-GlyProAlaG 985
QY 552 CAGATGAAGGCGCGCGCCATAGCTCCAGCCAGTGGTCTACTGCTGAGTGCCTAGG 493
Db 985 lySerValGlyProValGlyAlaValGly-ProArgGlyProSerGlyProGlnGlyLe 1004
QY 492 AGCGGAGACAGACAGCCCGC-----AGCAGTGGACCAATGCCAGCAGCATGTCATGAAC 436
Db 1005 ArgGlyAspLysGlyGluProGlyGlyLysGlyPro----- 1016
QY 435 TTCCTCTACCCCGCTTCCAGCAGCAGAGCGGCGCAGATAGTGTATGCTGCGCGCAAA 376
Db 1017 -----ArgGly-----LeuProGlyLeuLys 1023
```

QY 375 CACACCTCCAGGCAAGAGTTGACGAGTTCACACGAGAGAGTGGGCTTTCCGT---GC 319  
 Db 1024 GlyHisasn-Gly-----LeuGlnGlyLeuProGlyLeuAlaGlyGlnHisGlyAspG 1041  
 QY 318 CGCAGCAGGCGGTACCCACAGCCTCTGACCATAGTGGCGGCGGCGG----- 269  
 Db 1041 nGlyAlaProGlySerValGlyProAlaGlyPro-ArgGlyProAlaGlyProSerGlyP 1061  
 QY 268 --TAGGCTCAGGGCGGCTTACAGGACTCCAGAACTG-----CITCGTC 226  
 Db 1061 roAlaGlyLysAspGlyArgThrGlyGlnProGlyThrValGlyProAlaGlyIleArgG 1081  
 QY 225 TCGCTCTGCTCAGAGAGCTCGGCGCTCTCTCTGCTGCTCGCGCCCAACTGCTAGGAATC 166  
 Db 1081 lySerGlnGlySerGlnGlyProAlaGlyProGly-----ProProGlyP 1097  
 QY 165 AGCCAGGCGCCCAATTTCTGCGAGCCCTTTGGTGCCCGT----- 128  
 Db 1097 roProGlyProPro-----GlyProSerGlyGlyGlyTyAspPheGlyTyGluGlyA 1115  
 QY 127 -----CCAGCTTCTCAGCCCATGCTC 107  
 Db 1115 spPheTyArgAlaAspGlnProArgSerProProSerLeu 1128

RESULT 23  
 CA17\_HUMAN  
 ID CA17\_HUMAN STANDARD; PRT: 2944 AA.  
 AC Q02388; Q14054; Q16507;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC collagen).  
 GN COL7A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94327588; PubMed-8051117;  
 RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;  
 RT "Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII) chain and identification of intragenic polymorphisms.";  
 RL J. Biol. Chem. 269:20256-20262(1994).  
 RN [2]  
 RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE-93338437; PubMed-1307247;  
 RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G., Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E., Uitto J.;  
 RT "The large non-collagenous domain (NC-1) of type VII collagen is amino-terminal and chimeric. Homology to cartilage matrix protein, the type III domains of fibronectin and the A domains of von Willebrand factor.";  
 RL Hum. Mol. Genet. 4:475-481(1995).  
 RN [3]  
 RP SEQUENCE OF 815-1439 FROM N.A.  
 RX MEDLINE-91334380; PubMed-1871109;  
 RA Parente M.G., Chung L.C., Ryyanen J., Woodley D.T., Wynn K.W., Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;  
 RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).  
 RN [4]  
 RP SEQUENCE OF 369-1255 FROM N.A.  
 RX MEDLINE-93107742; PubMed-1469284;  
 RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisanan P.S., Cook M.E., Wright J., Brigaman R.A., Hunt S.W. III;  
 RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion proteins involved in tissue-specific organization of extracellular matrix.";

RL J. Invest. Dermatol. 99:691-696(1992).  
 RN [5]  
 RP SEQUENCE OF 340-675 FROM N.A.  
 RX MEDLINE-92231902; PubMed-1567409;  
 RA Tanaka T., Takahashi K., Furukawa F., Inamura S.;  
 RT "Molecular cloning and characterization of type VII collagen cDNA.";  
 RL Biochem. Biophys. Res. Commun. 183:958-963(1992).  
 RN [6]  
 RP SEQUENCE OF 2395-2944 FROM N.A.  
 RX MEDLINE-93271985; PubMed-8499916;  
 RA Greenspan D.S.;  
 RT "The carboxyl-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene.";  
 RL Hum. Mol. Genet. 2:273-278(1993).  
 RN [7]  
 RP SEQUENCE OF 1-87 FROM N.A.  
 RX MEDLINE-94375010; PubMed-8088784;  
 RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W., Uitto J., Greenspan D.S.;  
 RT "Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene.";  
 RL Genomics 21:169-179(1994).  
 RN [8]  
 RP REVIEW ON DEB VARIANTS.  
 RX MEDLINE-98041696; PubMed-9375848;  
 RA Jaervikallio A., Pulkkinen L., Uitto J.;  
 RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in the type VII collagen gene (COL7A1).";  
 RL Hum. Mutat. 10:338-347(1997).  
 RN [9]  
 RP VARIANT RDEB LYS-2798.  
 RX MEDLINE-93291877; PubMed-8513326;  
 RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y., Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;  
 RT "A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa.";  
 RL Nat. Genet. 4:62-66(1993).  
 RN [10]  
 RP VARIANT DDEB SER-2040.  
 RX MEDLINE-94224777; PubMed-8170945;  
 RA Christiano A.M., Ryyanen M., Uitto J.;  
 RT "Dominant dystrophic epidermolysis bullosa: identification of a Gly-->Ser substitution in the triple-helical domain of type VII collagen.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).  
 RN [11]  
 RP VARIANT PEB-DDEB CYS-2623.  
 RX MEDLINE-96081220; PubMed-8541842;  
 RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;  
 RT "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and identification of a glycine-to-cysteine substitution in the triple-helical domain of type VII collagen.";  
 RL Hum. Mol. Genet. 4:1579-1583(1995).  
 RN [12]  
 RP VARIANT DDEB ARG-2043.  
 RX MEDLINE-95164985; PubMed-7861014;  
 RA Christiano A.M., Morricone A., Paradisi M., Angelo C., Mazzanti C., Cavallieri R., Uitto J.;  
 RT "A glycine-to-arginine substitution in the triple-helical domain of type VII collagen in a family with dominant dystrophic epidermolysis bullosa.";  
 RL J. Invest. Dermatol. 104:438-440(1995).  
 RN [13]  
 RP VARIANTS RDEB AND DDEB.  
 RX MEDLINE-96220218; PubMed-8644729;  
 RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;  
 RT "Glycine substitutions in the triple-helical region of type VII collagen result in a spectrum of dystrophic epidermolysis bullosa phenotypes and patterns of inheritance.";  
 RL Am. J. Hum. Genet. 58:671-681(1996).



Db 1349 ProGlyAlaProGlyGlnValIleGlyGlyGluGlyProGlyLeuProGlyArgLysGly 1368  
Qy 1745 TGCCCGGCGCCGAA---CCACCTCGCTCGGTGGCTCACCCACCACACACGTACGG 1689  
Db 1369 AspProGlyProSerGlyProGlyProArg-Gly- 1383  
Qy 1688 AGACATCACAGGAGGCGCC---GCAGCGCGG 1656  
Db 1383 YAspProGlyProArgGlyProGlyLeuProGlyThrAlaMetLysGlyAspLysGly 1403  
Qy 1655 GTGAGGTGGAGACAGCCACTCCCTCCAGCACCCAGCTGTCTATTAGGAGAGGAGCTC 1596  
Db 1403 YAspArg-GlyGluArgGly---ProProGlyPro- 1416  
Qy 1595 CAGGCTTAGGCGCTGCAGAGAACTGTCTCATCAGGCTGTCTCACTGCTAGCACCTCCAG 1536  
Db 1416 LyGlyIleAlaProGly- 1421  
Qy 1535 TGTCCTCCCTCGTATTGGGAGGAAACACTGCTTCTCCGGTGTGTAGAGGAGGCGCAGTG 1476  
Db 1422 --GluPro-GlyLeu---ProGlyLeuProGlySerProGlyProGlyGly---ProVal 1438  
Qy 1475 TGTAGGCGAGACTCGAGGCTGAGAGGTGAACCGGTGAGGCGGCTGAGCTGTCA 1416  
Db 1439 GlyProGlyLysLysGlyGlyLysGlyAspSerGluAspGlyAlaProGlyLeuPro 1458  
Qy 1415 CCAGCGCCACACTGTGGGACAGGCTGTGCACCGCGGACAGCCAGGAAAGCTG- 1362  
Db 1459 GlyGlnProGlySerProGlyGluGlnGlyProArgGlyProProGlyAlaIleGlyPro 1478  
Qy 1361 -----CCACACTGGCCAAATAGACTCTCGAGTGGCGAATCGCTGCACAGCCGCT 1311  
Db 1479 LysGlyAspArgGlyPheProGlyProLeuGlyGluAlaGlyGlyGlyGlyGly 1498  
Qy 1310 CCATGACAGAGAGAACAGGAGATGCGGCACTGCAGGAACAGCCCGGCTGCCCA 1251  
Db 1499 Pro-----ProGly-----ProAlaGlySerArg---GlyLeuPro 1509  
Qy 1250 TCCGAAGCGCTTCATAGTGTCTCCGGCTCGGTGCGCGGCTCAGCTCTGGGACGC 1191  
Db 1510 GlyValAlaGlyArgProGlyAlaLysGlyProGlyGlyProGlyProThrGlyArg 1529  
Qy 1190 CCTGGTACAGCCCTCGCCAGCAATTCGTGTAAACAGCTGAAGGTGATGAGTCCCA 1131  
Db 1530 GlnGlyGlyGlyGlyGlyGlyProGlyArgProGlyAspProAla-----ValVal 1545  
Qy 1130 TCCAGCTGCACAGTGCACGAGAGAGCGCGGCGAGGTGCGGGCATCGGC- 1077  
Db 1546 GlyProAlaValAlaGlyProLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1565  
Qy 1076 AGCAGCTGTGTGCGCGGAGAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1017  
Db 1566 GlyAlaThrGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1585  
Qy 1016 GGCATGACAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 957  
Db 1586 GlyProLysGlyAspProGlyAsp---ArgGlyProIleGly---LeuThrGlyArgAla 1603  
Qy 956 GGCCAGCGCTGCCT-----ProGlyAspSerGlyProGlyGlyGlyGlyGlyGlyGly 939  
Db 1604 GlyPro-----ProGlyAspSerGlyProGlyGlyGlyGlyGlyGlyGlyGlyGly 1621  
Qy 938 CAGCCACAGCAGTGTGCTGCTACGAGGTGAGAGATGAGGCTGAGGAGGCGGCGGCGG 879  
Db 1622 GlyProGlyProValGlyProArg-----GlyArgGlyGlyGlyGlyGlyGlyGly 1639  
Qy 878 GGCAGT-----CCTCTGGGTGGCGGCT----- 855  
Db 1640 GlyAspGlyGlyProGlyAspProGlyLeuProGlyLysAlaGlyGlyGlyGlyGly 1659  
Qy 854 -----AGGGGCGCA----- 828  
Db 1660 ArgGlyAlaProGlyValArgGlyProValGlyGlyGlyGlyGlyGlyGlyGlyGly 1679

Qy 827 CAATGGCAGGAGGTAGCCAGGAGAGCCCAAGACTGATCATGAAGCATAGACAG 768  
Db 1680 GluAspGlyArgAsnGlySerProGlySerGlyPro-----LysGlyAspArg 1696  
Qy 767 AGTAGGCTGCGACAGTGGTCCGGGTCCCGAAGAGAGTACAGAGAGCGGCTCCAGTG 708  
Db 1697 GlyGluProGlyProProGlyPro-----ProGlyArgLeuValAspThrGlyProGlyAla 1715  
Qy 707 ---GAGTGAAGCACACCTGGCCACAGAGTCCAGCAGCGCCCGCCAGC- 666  
Db 1716 ArgGlyLysGlyGluProGlyAspArgGlyGlnGlyProArgGlyProLysGlyAsp 1735  
Qy 665 CCAGGATGAGCAGTCCAGCTCCAGGCGCTGGATCCGGCAGCAGCAGCAGCCTCTAGCC 606  
Db 1736 ProGly-----LeuProGlyAla-----ProGly- 1743  
Qy 605 AGCCGCGCTTGGGTATGAAGAGAGGCTCAGCAGGATGC-----CCAGG 561  
Db 1744 -----GluArgGlyIleGlyGlyPheArgGlyProPheGlyProGlyProGln 1757  
Qy 560 ACAGTCCCGCAGATGAAGGCGCGGCGGCGCATAGCTCCAGCGGCTGCTGCTGGTG 501  
Db 1758 GlyAspProGlyValArgGlyProAlaGlyGlyLys----- 1769  
Qy 500 AGCTTAGGAGCGGACACAGCAGCAGCAGCAGCTGGACCAATGCCAGCAGCAGTGTCA 441  
Db 1770 -----GlyAspArgGlyProProGlyLeuAspGly----- 1779  
Qy 440 TGAACCTCTCTCTACCCCTCCAGCAGCAGGCGGCGCATAGTGTGATGCTGCTGG 381  
Db 1780 ---ArgSerGlyLeuAspGlyLysProGlyAlaAlaGly----- 1791  
Qy 380 CCAACACACACTCCAGGCGCAAGG---TTAGCAGGTGACAGCAGAGCTGGGCTTTC 324  
Db 1792 ProSerGlyProAsnGlyAlaGlyLysAlaGlyAspProGlyArgAspGlyLeuPro 1811  
Qy 323 GGT---GCCGACAGGCGGCTCAGCAGCAGCAGCTCTGACCATAGTGG- 278  
Db 1812 GlyLeuArgGlyGluGlnGlyLeuProGlyProSerGlyPro-ProGlyLeuProGlyLys 1831  
Qy 277 -----CCAGGCGGTAGGCTCAGGCGGCGCTTCAGGCGCTCCA- 239  
Db 1831 sProGlyGluAspGlyLysProGlyLeuAsnGlyLysAsnGlyGlyAspProGly 1851  
Qy 238 -GAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
Db 1851 yGluAspGlyArgLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1871  
Qy 179 ACTGCTAGTAATCAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 123  
Db 1871 pGlyProLysGlyGluArgGlyAlaProGlyIleLeuGlyProGlnGly-ProProGlyL 1891  
Qy 122 TTCTCAGCCCATGCTCAACACCTGCT- -GCTGTGGGCGACCTCAGTGGGCGGCTCTC 66  
Db 1891 euProGlyProValGlyProProGlyGlnGlyPheProGlyValProGlyGlyThrGly 1911  
Qy 65 ATCAGTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 45  
Db 1911 rolyGlyAspArgGlyGluThrGlySerLysGlyGlnGlyLeuProGlyGluArg 1931  
Qy 44 GCGCGCGGTGTCACCGGAGCC 22  
Db 1931 lyLeuArgGlyGluProGlySer 1938

## RESULT 24

-ID CA24\_ASCSU STANDARD; PRT; 1763 AA.  
AC P27393;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Collagen alpha 2(IV) chain precursor.







Db 1266 lnProGlyLysSerIleThrGlyProLysGlyAsnAlaGly-LeuProGlyLeuProGly 1285  
QY 2824 AAATGCTGTACCAAGGTTAGGCTGTGAAGGAGGTAGAGGT----- 2868  
Db 1286 LysAspGlyLeuProGlyLeuProGlyLeuLysGlyGluProGlyLysProGlyTyrAla 1305  
QY 2869 -----GGGGCT 2874  
Db 1306 GlyAlaAlaGlyIleLysGlyGluProGlyLeuProGlyIleProGlyAlaLysGlyGlu 1325  
QY 2875 TCAGTCTCAAGGCTCCCTAACCCAC-----CCTCTTCTCTGGC 2916  
Db 1326 ProGlyLeuSerGlyIleProGlyLysArgGlyAsnAspGlyIleProGlyLysProGly 1345  
QY 2917 CCAGCTGGTCCGCCACATCCA-----CTCCCTCTACTCTC 2955  
Db 1346 ProAlaGlyLeuProGlyLeuProGlyMetLysGlyGluSerGlyLeuProGlyProGln 1365  
QY 2956 TCTAGGACTGGCTGATGAAGGCACTGCCCAAAATTTCCCTACCCCAACTTCCC 3012  
Db 1366 GlyProAlaGlyLeu---ProGlyLeuProGlyLeuLysGlyGluProGlyLeuPro 1383  
RESULT 25-  
CA25\_HUMAN STANDARD; PRT; 1496 AA.  
AC P05997;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Collagen alpha 2(V) chain precursor.  
GN COL5A2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE OF 1-463 FROM N.A.  
RX MEDLINE=89123368; PubMed=2914927;  
RA Woodbury D., Benson-Chanda V., Ramirez F.;  
RT "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms  
RT to the structural criteria of a fibrillar procollagen molecule.";  
RL J. Biol. Chem. 264:2735-2738(1989).  
RN [2]  
RP SEQUENCE OF 398-1496 FROM N.A.  
RX MEDLINE=87146331; PubMed=3029669;  
RA Weil D., Bernard M.P., Gargano S., Ramirez F.;  
RT "The pro alpha 2(V) collagen gene is evolutionarily related to the  
RT major fibrillar-forming collagens";  
RL Nucleic Acids Res. 15:181-198(1987).  
RN [3]  
RP SEQUENCE OF 1227-1496 FROM N.A.  
RX MEDLINE=85289337; PubMed=2411731;  
RA Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;  
RT "Complete primary structure of the human alpha 2 type V procollagen  
RT COOH-terminal propeptide.";  
RL J. Biol. Chem. 260:11216-11222(1985).  
RN [4]  
RP SEQUENCE OF 1449-1496 FROM N.A.  
RX MEDLINE=89138450; PubMed=3224983;  
RA Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,  
RA Ramirez F.;  
RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,  
RT located on the long arm of human chromosome 2.";  
RL Genomics 3:275-277(1988).  
RN [5]  
RP SEQUENCE OF 208-227.  
RX TISSUE=Placenta;  
RA Mann K.;  
RT "Isolation of the alpha 3-chain of human type V collagen and  
RT characterization by partial sequencing.";  
RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).  
RN [6]

RP SEQUENCE OF 288-297 AND 606-617.  
RC TISSUE=Bone;  
RX MEDLINE=94237164; PubMed=8181482;  
RA Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champiaud M.F.,  
RA Bouthillon M.M., Bernillon J., Wallach J.M., van der Rest M.;  
RT "Diversity in the processing events at the N-terminus of type-V  
RT collagen.";  
RL Eur. J. Biochem. 221:987-995(1994).  
RN [7]  
RP DISEASE.  
RX PubMed=9425231;  
RA Michalickova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;  
RT "Mutations of the alpha2(V) chain of type V collagen impair matrix  
RT assembly and produce Ehlers-Danlos syndrome type I.";  
RL Hum. Mol. Genet. 7:249-255(1998).  
RN [8]  
RP VARIANT EDS-II ARG-960.  
RX PubMed=9783710;  
RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,  
RA Burrows N.P.;  
RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type  
RT II.";  
RL J. Med. Genet. 35:846-848(1998).  
CC -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
CC (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE  
CC COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS  
CC TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.  
CC -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN  
CC MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND  
CC ONE ALPHA 3(V) CHAINS IN PLACENTA.  
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome  
CC gravis or type I (EDS-I), and Ehlers-Danlos syndrome mitis or type  
CC II (EDS-II). Ehlers-Danlos syndrome is a genetically and  
CC phenotypically heterogeneous connective-tissue disorder  
CC characterized by loose-jointedness and fragile, velvety,  
CC stretchable, bruisable skin that heals with peculiar 'fingertette-  
CC paper' scars. EDS-I and EDS-II are autosomal dominant traits.  
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J04478; AAA51859.1; -  
DR EMBL: X04758; CAA28454.1; -  
DR EMBL: M11718; AAA52058.1; -  
DR PIR: A25374; A25374.  
DR PIR: A25874; A25874.  
DR PIR: A30017; A30017.  
DR PIR: A31427; A31427.  
DR Genew: HGNC:2210; COL5A2.  
DR MIM: 120190; -  
DR MIM: 130000; -  
DR MIM: 130010; -  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR000885; Fib.collagen\_C.  
DR InterPro: IPR001007; VWFC.  
DR Pfam: PF00093; vwc; 1.  
DR Pfam: PF01391; Collagen; 18.  
DR Pfam: PF01410; COLFI; 1.  
DR ProDom: PD000007; Collagen; 5.  
DR ProDom: PD002078; Fib.collagen\_C; 1.  
DR SMART: SM00038; COLFI; 1.  
DR SMART: SM00214; VWC; 1.  
DR PROSITE: PS01208; VWFC; 1.  
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;

KW Disease mutation. 26  
 FT SIGNAL 1 1226  
 FT CHAIN 27 1226  
 FT PROPEP 1227 1496  
 FT DOMAIN 39 97  
 FT MOD\_RES 290 290  
 FT MOD\_RES 293 293  
 FT MOD\_RES 296 296  
 FT MOD\_RES 608 608  
 FT MOD\_RES 614 614  
 FT MOD\_RES 960 960  
 FT VARIANT 960 960  
 FT CONFLICT 292 292  
 FT CONFLICT 1418 1418  
 FT CONFLICT 1438 1438  
 FT CONFLICT 1460 1460  
 FT CONFLICT 1496 1496  
 SQ SEQUENCE 1496 AA: 144720 MW: 82827C17A8644F5A CRC64;

Alignment Scores:  
 Pred. No.: 2.18e-07 Length: 1496  
 Score: 312.00 Matches: 226  
 Percent Similarity: 30.80% Conservative: 37  
 Best Local Similarity: 26.46% Mismatches: 255  
 Query Match: 4.86% Indels: 336  
 DB: 1 Gaps: 51

US-09-759-143-110 (1-3410) x CA25\_HUMAN (1-1496)

QY 26 CCGGGTGC-----AGCGCGCGCCTCGG-----CCAGGA 55  
 DB 242 ProGlyAspProGlyProMetGlyProLeuGlySerArgGlyProGlyProGlyProGly 261  
 QY 56 TCTGAGTGATGAGAGCGTGTCCACACGAGG-----85  
 DB 262 LysProGlyGluAspGlyGluProGlyArgAsnGlyAsnProGlyGluValGlyPheAla 281  
 QY 86 -----TGCACACAGCAGCAGGTGTGAGCATGGGTGAGTGAAGC 124  
 DB 282 GlySerProGlyAlaArgGlyPheProGlyAlaProGlyLeuPro---GlyLeuLysGly 300  
 QY 125 TGGACCGCCACCAAGGCGTG-----145  
 DB 301 HisArgGlyHisLysGlyLeuGlyProLysGlyGluValGlyAlaProGlySerLys 320  
 QY 146 -----GCAGAAATGGCGCCTGG-----CTGATTCCTAGGCAGTTG 181  
 DB 321 GlyGluAlaGlyProThrGlyProMetGlyAlaMetGlyProLeuGlyProArgGlyMet 340  
 QY 182 CGGCACCAAGGAGGAGCGCGCTTCTGGAGCAGACCGCAGAGCAGGAGTCTCGG 241  
 DB 341 ProGlyGluArg-GlyArgLeuGlyProGlnGlyAlaProGlyGlnArgGlyAlaHisGly 360  
 QY 242 AGTCCCTGAAGCGCC-----257  
 DB 360 yMetProGlyLysProGlyProMetGlyProLeuGlyIleProGlySerSerGlyPhePr 380  
 QY 257 -----257  
 DB 380 oGlyAsnProGlyMetLysGlyGluAlaGlyProThrGlyAlaArgGlyProGlyGlyPr 400  
 QY 258 -----CCTGAGCCCTACCGCGCTGCCCACTATGCT-----CCAG 292  
 DB 400 oGlnGlyGlnArgGlyGluThrGlyProGlyProGlyProVal-GlySerProGlyLeuProG 420  
 QY 293 AGGCTGTGGGTGAGCGCCCTGCTGCGCACCGCAA-----AGCCCAAGCTCTTG 340  
 DB 420 lyAlaIleGlyThrAspGlyThrProGlyProLysGlyProThrGlySerProGlyThrS 440  
 QY 341 CTGCTCAACCTGTCAACCTTTGGCGTGGAGGTGTGTTGGCCGACGATCACCCTATG 400  
 DB 440 erGlyProGlySerAlaGlyProProGlySerProGlyProGln-----455

QY 401 CCGCCTCTGCTGCTGGAAAGTGGGGGTAGAGAGAGAAAGTTTCATGACCATGGTCTGGGCATT 460  
 DB 456 --GlySerThrGlyProGlnGlyAsnSerGlyLeuProGlyAsp-----ProGlyPheL 473  
 QY 461 GGTCCAGTGTGGCGCTGTGTGTCCTCCGCTCAGTCTAGGCTCAGCAGCAGTACACTGGCG - 519  
 DB 473 ysGlyGluAlaGlyProLysGlyGluProGlyProHisGlyIleGlnGlyProIleGlyLp 493  
 QY 520 -----TGCACCTATGGCCGCGC - 537  
 DB 493 roProGlyGluGlyLysArgGlyProArgGlyAspProGlyThrLeuGlyProProG 513  
 QY 538 ---CCGGCCCTTCACTCTGGGCACTGTC-----CTTGGGCATCCF---GCTGAGCCTCTTT 586  
 DB 513 lyProValGlyGluArgGlyAlaProGlyAsnArgGlyPheProGlySerAspGlyLeuP 533  
 QY 587 CTCATCCCAAGGCGCGCTGTGCTAGCAGGCTGTGTCGCGGATCCCGAGCCCTCGAG 646  
 DB 533 roGlyProLysGlyAlaGlnGlyGluArgGlyProVal---GlySerSerGlyProLysG 552  
 QY 647 CTGGCACTGCTCATCTCTGGGCGT---GGGGTGTGTGGACTTCTGTGGCCAGGTGTGCTTC 703  
 DB 552 lySerGlnGlyAspProGlyArgProGlyGluProGlyLeu-----ProGlyAlaArg 570  
 QY 704 ACTCCACTGGAGGCCCTGCT-----CTCTGACCTCTTCCGGGAC 742  
 DB 570 lyLeuThrGlyAsnProGlyValGlnGlyProGlyLysLeuGlyProLeuGlyAlaP 590  
 QY 743 CCGGA-----CCACTGTCCGACGCTACTCTGTCTATGCTTCTCATGATCAGT 790  
 DB 590 roGlyGluAspGlyArgProGlyProProGly-----S 601  
 QY 791 CTGTTGGGCTGCTGGGTACTCTCTGCTGCCATTCAGTGGGACACAGTGCCTGGCC 850  
 DB 601 erIleIleLysGlyGlnProGlyThr-----MetGly-----LeuProGlyP 616  
 QY 851 CCCTA-----CCTGGCACCCAGGAGGAGTCCCTTTTGGCCTGCTCACC 895  
 DB 616 roLysGlySerAsnGlyAspProGlyLysProGlyGluAla-----GlyAsnP 632  
 QY 896 CTCATCTCTCCACCTCGGTAGCAGCCACACTGTGTGGTGGAGGAGGAGCCTGGCC 955  
 DB 632 roGlyValProGlyGlnArgGly-----AlaProGlyLysAspGlyLysValGlyP 649  
 QY 956 CCCAC-----CGAGCCAGCAGAGGCTCTCGGCCCTCC 991  
 DB 649 roTyrglyProProGlyProProGlyLeuArgGlyGluArgGlyGlnGlyPro----- 667  
 QY 992 TTGTCGCCCACTGTGTCTCCATGCGCGCGCTTGGCTTTCGGGAACCTGGCGCCCTG 1051  
 DB 668 -----ProGlyProThrGlyPhe-----GlnGlyHisProG 678  
 QY 1052 CTTCCCGGCTGCACAGCTGTGTGTCGCGCATGCCCGCACCCCTGCGCGGCTCTTCGTG 1111  
 DB 678 lyProGlyProProGly----- 684  
 QY 1112 GCTGAGCTGTGCAGCTGTGATGCACTCATGACCTTCACGCTGTTTACACGATTTCTGT 1171  
 DB 685 -----GluGly-----GlyLysProG 690  
 QY 1172 GCGAGGGGCTGTACAGGGCGTCCAGAGAGTGTAGCGCGGCGCACCGCGCGGAGACAC 1231  
 DB 690 lyAspGlnGlyValProGlyGlyProGlyAla---ValGlyProLeuGlyPro----- 706  
 QY 1232 TATGATGAAGCGTTCGGATGGCAGCCTGG-----GCTGTTCTCGATGCGGCCATC 1285  
 DB 707 -----ArgGlyGluArgGlyAsnProGlyGluArgGlyGluProGlyIleThrGlyL 724  
 QY 1286 TCCCT-----GGTCTTCTCTGTCATGGACCGCTGGTGCACGCTATCGGCCT 1336  
 DB 724 euProGlyGluLysGlyMetAlaGlyHisGlyProAspGly----- 738  
 QY 1337 CGAGCAGTCTATTGGCCAGTGTGCGAGCTTTTCCCTGTGGTGTGCGGTCGCCACATGCTG 1396

Db 739 -----ProLysGly-----SerProG 744  
 QY 1397 TCCACAGTGTGGCGTGTGACACCTTACCGCCCTCACCGGGTTCACCTTCACGCC 1456  
 Db 744 LysProSerGlyThrProGlyAspThr----- 752  
 QY 1457 CTGCAGATCTCGCCCTACACACTGCCCTCCCTACACCGCGGAGACAGGTCTTCTG 1516  
 Db 753 --GlyProProGlyLeuGln---GlyMetProGlyGluArgGlyLleAlaGlyThrProG 771  
 QY 1517 CCCAATACCGAGGACACTGGAGGTCTAGCAGTGGAGACACCTGATGACACGCG--- 1573  
 Db 771 LysPro---LysGlyAspArgGlyClyIleGlyGlyLysGlyAlaGlyThrAlaGly 789  
 QY 1574 -----TTCTCCAGCGCCCTACAGCTGGAGCTCCCTCCCTCCCTATGGACAC 1618  
 Db 790 AsnAspGlyAlaGlyGlyLeuProGlyProGlyLeuValGlyPheAlaGlyLeu 808  
 QY 1619 GTGGGTGCTGGAGCAGT-----GGCCTGCTCCACCTCCA----- 1654  
 Db 809 LeuGlyGlyLysGlyGluProGlyProArgGlyLeuValGlyProGlySerArgGly 828  
 QY 1655 ---CCCGCGCTCTCGGGCGCTCTGCTGTGATCTCCGTACGTGTG----- 1699  
 Db 829 AsnProGlySerArgGlyGluAsnGlyProThrGlyAlaValGlyPheAlaGlyProGln 848  
 QY 1700 -----GTGGTGGGTGAGCCACCGAGCGCGGTGGTTCGGGC 1738  
 Db 849 GlySerAspGlyGlnProGlyValLysGlyGluProGlyGluProGlyGlnLysGly 868  
 QY 1739 CGGGCATCTGCGGACCTGCG-----CATCTGGATAGTCTCTCTGCTG 1786  
 Db 868 spAlaGlySerProGlyProGlnGlyLeuAlaGlySerProGly----- 882  
 QY 1787 TCCAGGTGGCCCATCTCTGTATGGGTCCAT-----TGTCAGCTCAGC 1834  
 Db 883 --ProHisGlyProAsnGlyValProGlyLeuLysGlyGlyArgGlyThrGlnGlyPro 902  
 QY 1835 CAGTCTGCTACTGCTATATGCTCTGCGCGAGCGTGGTCTGTCGCGCATTTACTTT 1894  
 Db 902 ro-----GlyAlaThrGlyPheProGlySerAla----- 911  
 QY 1895 GCTACACAGGTAGTATTGTGACAGAGCGACTTGGCCAAATACTACGGTAGAAAACTTC 1954  
 Db 911 ----- 911  
 QY 1955 AGCATATTGGGTGGAGGCGCTGCTACCTGCTCCAGCTCCCGCTCTCTGTAGCCGC 2014  
 Db 912 -----GlyArgValGlyProProGlyProAlaGlyAlaProGlyProAlaGlyPro 928  
 QY 2015 ATGGGGTGGCGGCG-----TGGCGGCCA 2038  
 Db 929 LeuGlyGluProGlyLysGlyGluGlyProPro 938  
 RESULT 26  
 CA21\_CHICK STANDARD; PRT; 1362 AA.  
 AC P02467; Q90795; Q90797; P87492; Q90758; Q92014; P87491; Q90792;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(I) chain precursor (Fragments).  
 GN ColIA2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus  
 OX NCBI-TaxID=9031;  
 RN [1]  
 RP SEQUENCE OF 1-245; 262-448 AND 466-1362 FROM N.A.  
 RX MEDLINE=86185168; PubMed=3868961;  
 RA Boedtker H., Finer M., Aho S.,

"The structure of the chicken alpha 2 collagen gene."; Ann. N.Y. Acad. Sci. 460:85-116(1985).  
 [2]  
 RN SEQUENCE OF 1-89 FROM N.A.  
 RX MEDLINE=83246518; PubMed=6135195;  
 RA Tate V.E., Finer M.H., Boedtker H., Doty P.;  
 RT "Chick pro alpha 2(I) collagen gene: exon location and coding potential for the prepropeptide."; Nucleic Acids Res. 11:91-104(1983).  
 [3]  
 RN SEQUENCE OF 1-14 FROM N.A.  
 RX MEDLINE=82060240; PubMed=6946474;  
 RA Vogel G., Ohkubo H., Sobel M.E., Yamada Y., Pastan I., de Crombrughe B.;  
 RT "Structure of the promoter for chicken alpha 2 type I collagen gene."; Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).  
 [4]  
 RN SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=84297217; PubMed=6473103;  
 RA Aho S., Tate V.E., Boedtker H.;  
 RT "Location of the 11 bp exon in the chicken pro alpha 2(I) collagen gene."; Nucleic Acids Res. 12:6117-6125(1984).  
 [5]  
 RN SEQUENCE OF 1-79 FROM N.A.  
 RX MEDLINE=88056316; PubMed=3678834;  
 RA Finer M.H., Boedtker H., Doty P.;  
 RT "Construction and characterization of cDNA clones encoding the 5' end of the chicken pro alpha 1(I) collagen mRNA."; Gene 56:71-78(1987).  
 [6]  
 RN SEQUENCE OF 78-92.  
 RC TISSUE=Skin;  
 RX MEDLINE=71115216; PubMed=5544653;  
 RA Highberger J.H., Kang A.H., Gross J.;  
 RT "Comparative studies on the amino acid sequence of the alpha 2-CB2 peptides from chick and rat skin collagens."; Biochemistry 10:610-616(1971).  
 [7]  
 RN SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.  
 RX MEDLINE=82058081; PubMed=6272119;  
 RA Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.;  
 RT "Structure of the pro alpha 2 (I) collagen gene."; Nature 294:129-135(1981).  
 [8]  
 RN SEQUENCE OF 78-92.  
 RC TISSUE=Skin;  
 RX MEDLINE=70131186; PubMed=4313735;  
 RA Kang A.H., Gross J.;  
 RT "Amino acid sequence of cyanogen bromide peptides from the amino-terminal region of chick skin collagen."; Biochemistry 9:796-804(1970).  
 [9]  
 RN SEQUENCE OF 78-92 AND 415-448.  
 RC TISSUE=Skin;  
 RX MEDLINE=69285369; PubMed=5809220;  
 RA Kang A.H., Igarashi S., Gross J.;  
 RT "Characterization of the cyanogen bromide peptides from the alpha-2 chain of chick skin collagen."; Biochemistry 8:3200-3204(1969).  
 [10]  
 RN SEQUENCE OF 78-92 AND 415-448.  
 RC TISSUE=Bone;  
 RX MEDLINE=69206882; PubMed=5785233;  
 RA Lane J.M., Miller E.J.;  
 RT "Isolation and characterization of the peptides derived from the alpha 2 chain of chick bone collagen after cyanogen bromide cleavage."; Biochemistry 8:2134-2139(1969).  
 [11]  
 RN SEQUENCE OF 566-587 FROM N.A.  
 RX MEDLINE=79074829; PubMed=364479;  
 RA Lehrach H., Frischauf A.-M., Hanahan D., Wozney J., Fuller F.,

RA Crkvenjakov R., Boedtker H., Doty P.;  
 RT "Construction and characterization of a 2.5-kilobase procollagen  
 RL clone.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421(1978).  
 RN [12]  
 RN SEQUENCE OF 902-1362 FROM N.A.  
 RX MEDLINE=81160715; PubMed=6927845;  
 RA Fuller F., Boedtker H.;  
 RT "Sequence determination and analysis of the 3' region of chicken pro-  
 RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids  
 RT including the carboxy-terminal propeptide sequences.";  
 RL Biochemistry 20:996-1006(1981).  
 RN [13]  
 RN SEQUENCE OF 998-1169 AND 1234-1362 FROM N.A.  
 RX MEDLINE=81264246; PubMed=6267043;  
 RA Dickson L.A., Nimmiya Y., Bernard M.P., Pesciotta D.M., Parsons J.,  
 RA Green G., Eikenberry E.F., de Crombrughe B., Vogeli G., Pastan I.,  
 RA Fietzek P.P., Olsen B.R.;  
 RT "The exon/intron structure of the 3'-region of the pro alpha 2(I)  
 RT collagen gene.";  
 RL J. Biol. Chem. 256:8407-8415(1981).  
 RN [14]  
 RN SEQUENCE OF 932-954 AND 968-980 FROM N.A.  
 RX MEDLINE=81064671; PubMed=6159982;  
 RA Avvedimento V.E., Vogeli G., Yamada Y., Maizel J.V. Jr., Pastan I.,  
 RA de Crombrughe B.;  
 RT "Correlation between splicing sites within an intron and their  
 RT sequence complementarity with U1 RNA.";  
 RL Cell 21:689-696(1980).  
 RN [15]  
 RN SEQUENCE OF 126-161; 467-517 AND 926-954 FROM N.A.  
 RX MEDLINE=81112157; PubMed=7460017;  
 RA Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,  
 RA Irani M., Pastan I., de Crombrughe B.;  
 RT "The collagen gene: evidence for its evolutionary assembly by  
 RT amplification of a DNA segment containing an exon of 54 bp.";  
 RL Cell 22:887-892(1980).  
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC (FIBRILLAR FORMING COLLAGEN).  
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 CC HYDROXYAPATITE.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M25978; AAA69962.1; JOINED.  
 DR EMBL; M25979; AAA69962.1; JOINED.  
 DR EMBL; M25980; AAA69962.1; JOINED.  
 DR EMBL; M25981; AAA69962.1; JOINED.  
 DR EMBL; M25982; AAA69962.1; JOINED.  
 DR EMBL; M25983; AAA69962.1; JOINED.  
 DR EMBL; M25984; AAA69962.1; JOINED.  
 DR EMBL; J00821; AAA51611.1; JOINED.  
 DR EMBL; K00792; AAA51611.1; JOINED.  
 DR EMBL; J00830; AAA51613.1; JOINED.  
 DR EMBL; J00829; AAA51613.1; JOINED.  
 DR EMBL; J00837; AAA51614.1; JOINED.  
 DR EMBL; J00812; AAA51615.1; JOINED.  
 DR EMBL; J00811; AAA51615.1; JOINED.  
 DR EMBL; J00814; AAA51615.1; JOINED.  
 DR EMBL; J00815; AAA51615.1; JOINED.  
 DR EMBL; X02657; CAA26493.1; JOINED.  
 DR EMBL; K00794; NOT\_ANNOTATED\_CDS.  
 DR EMBL; X00390; CAA23688.1; JOINED.  
 DR EMBL; X00760; CAA25330.1; JOINED.  
 DR EMBL; M17608; AAA48673.1; JOINED.  
 DR EMBL; M10581; AAA48637.1; JOINED.  
 DR EMBL; M10540; AAA48638.1; JOINED.  
 DR EMBL; J00828; AAA51612.1; JOINED.  
 DR EMBL; J00827; AAA51612.1; JOINED.  
 DR EMBL; J00831; NOT\_ANNOTATED\_CDS.  
 DR EMBL; J00832; NOT\_ANNOTATED\_CDS.  
 DR EMBL; J00833; NOT\_ANNOTATED\_CDS.  
 DR EMBL; J00822; NOT\_ANNOTATED\_CDS.  
 DR PIR; A02868; CGCH2S.  
 DR PIR; S07354; S07354.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001670; Fe-ADH.  
 DR InterPro; IPR000885; Fib.collagen\_C.  
 DR InterPro; IPR001431; Peptidase\_M16.  
 DR InterPro; IPR000244; Ribosomal\_L9.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR InterPro; IPR000360; Transketolase.  
 Alignment Scores:  
 Pred. No.: 2,32e-07 Length: 1362  
 Score: 311.50 Matches: 264  
 Percent Similarity: 31.02% Conservative: 54  
 Best Local Similarity: 25.76% Mismatches: 310  
 Query Match: 5.00% Indels: 398  
 DB: 57 Gaps: 57

US-09-759-143-110 (1-3410) x CA21\_CHICK (1-1362)

QY 2110 GCTGTGCAGCTACGACACCTCAGCAGCAGAGGTGTCGACGAGCCAGCCACATTACTTTGGC 2051  
 Db 20 AlaThrSerGlnHisValSerGluAlaSerAlaGlyArgLysGlyProArg----- 36  
 QY 2050 AGCAACAGAAATGCGCGGCGGCGAGCCGCCATGGG---GCTAACAGGAGCGGGAG 1994  
 Db 37 GlyAspLysGlyProGlnGlyGluArgGlyProGlyProGlyProGlyProGlyProGlyGlu 56  
 QY 1993 CTGGGA---CCAGTCGAGCGAGCCGCT-----CCACCCCAATGTCTGCTGAAATTTCTA 1943  
 Db 57 AspGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyPheAla 76  
 QY 1942 CGCTGAGTATTGGCCCAAGTCGCTCTGTCAATACTACTCTGTGTGACCAAGGTAATGCG 1883  
 Db 77 -AlaGlnTyr-AspProSerLysAlaAlaAspPheGlyProGlyProMetGlyLeuMetG 96  
 QY 1882 GACCA-----GACCCAGCGCTGCGGCACAC 1856  
 Db 96 lyProArgGlyProProGlyAlaSerGlyProGlyProGlyProGlyPheGlnGlyValP 116  
 QY 1855 CATATGGCAGTCACAGACTGGCTGAGCTGGACAATGGAGCCATTAACAGGCGATGGCG 1796  
 Db 116 roGlyGluProGlyGluProGlyGlnThrGlyProGlnGlyProArgGlyProProGlyP 136

```

QY 1795 CACCTGGACACGAGGAGGCAC---TATCCAGGATGGCGAGGTCCAGGAGTCACCGG 1739
Db 136 roProGlyLysAlaGlyGluAspGlyHisProGly-----LysProGlyArg---ProG 153
QY 1738 GC-----CCGGAACAC 1727
Db 153 lyGluArgGlyValAlaGlyProGlyGlyAlaArgGlyPheProGlyThrProGlyPro- 172
QY 1726 CCTGGCTCGGTGGCTTCACCCACACACACGTCAGGACATCAGGACACA----- 1673
Db 173 ProGlyPheLysGlyIleArgGlyHisAsnGlyLeuAspGlyLeuThrGlyGlnProGly 192
QY 1673 ----- 1673
Db 193 AlaProGlyThrLysGlyGluProGlyAlaProGlyGluAsnGlyThrProGlyGlnPro 212
QY 1672 -----GGCCCCGAGAGCGC 1658
Db 213 GlyAlaArgGlyLeuProGlyGluArgGlyArgIleGlyAlaProGlyProAla-GlyAl 232
QY 1657 GGTTGGA-----GGTGGAGACAGGCGCTGCTCCAGCA----- 1624
Db 232 aaArgGlySerAspGlySerAlaGlyProThrGlyProAla***** 252
QY 1624 ----- 1624
Db 252 *****GlyGluIleGlyProAlaGlyAsnGluG 272
QY 1623 -CCACGTGTCCATTAGG----- 1606
Db 272 yProThrGlyProAlaGlyProArgGlyGluIleGlyLeuProGlySerSerGlyProVa 292
QY 1605 -----AAGGGAGCTCCAGGCTTAGGGCTGGCAGGAAGCTGGTCATCAG 1562
Db 292 lGlyProProGlyAsnProGlyAlaAsnGlyLeu---ProGlyAlaLys-GlyAlaAlaG 311
QY 1561 GCTGTCTCTACTCTAGCACCTCCAGTGTCCCTCGTATTGGGAGGAACACCTGCTT 1502
Db 311 lyLeuPro-----GlyVal---AlaGlyAlaProGlyL 321
QY 1501 CTCCCGTGGTAGAGG-----AGGCCAGTGTAGGCGAGGATGTCAG 1457
Db 321 euProGlyProArgGlyIleProGlyProProGlyProAlaGlyProSerGlyAlaArg 341
QY 1456 GGCTGAGAGGTGAACCCGGTGAGGCGGCTGAAGCTGTCCACCGGCCAC-----ACT 1403
Db 341 lyLeu-ValGlyGluProGlyProAlaGlyAlaLysGlyGlyGlyGlyGlyGlyGlu 360
QY 1402 GTGGACAGGCATGTGGCACCGCACCGCCAGGAAAGCTGCCACACTGGCCAAATAGAC 1343
Db 361 ProGlyAlaAlaGlyProProGlyProProGlyProSerGlyGlyGlyGlyLysArgGly 380
QY 1342 TGCTGAGTGCCTGAATC-----GCTGCACCGCCGGTCCA-----TGA 1305
Db 381 ---SerAsnGlyGluProGlySerAlaGlyProProGlyProAlaGlyLeuArgGlyGlu 399
QY 1304 CCAGAGAGAGA-----CCAGGAGATGGCCCATCTGCAGGAACA----- 1266
Db 400 ProGlySerArgGlyLeuProGlyAlaAspGlyArgAlaGlyValMetGlyProAlaGly 419
QY 1266 ----- 1266
Db 420 AsnArgGlyAlaSerGlyProValGlyAlaLysGlyProAsnGlyAspAlaGlyArgPro 439
QY 1265 -GGCCCCAGGCTGCCATCCGACCGCTTTCATCATGTGCTCC----- 1224
Db 440 GlyGluProGlyLeuMetGlyProArg***** 459
QY 1223 -----GGCCCTCGGTGC 1212
Db 460 *****GlyPheProGlyAlaAspGlyArgValGlyProIleGly 479
QY 1211 CCGGCTCAGCTCTGGGCGCGCCCTGGTACAGCC-----CCTCGCCACGAAATCCGTGT 1158

```

---

```

Db 480 ProAlaGlyAsnArgGlyGluProGlyAsnIleGlyPheProGlyProLysGlyPro--- 498
QY 1157 AAACACAGCGTGAAGGTTCATGAGTGCATCCAGCTGCACAGCTGCACAGGAGGAGCCGGC 1098
Db 499 -----ThrGlyGluProGlyLysProGly 506
QY 1098 ----- 1098
Db 507 GluLysGlyAsnValGlyLeuAlaGlyProArgGlyAlaProGlyProGluGlyAsnAsn 526
QY 1097 -----CCAGGGTGGGGCGGCGGCGGAGCAGCACA----- 1071
Db 527 GlyAlaGlnGlyProProGlyValThrGlyAsnGlnGlyAlaLysGlyGluThrGlyPro 546
QY 1070 -----GCTGTGTGACGCGGGGAGGAGCAGG 1047
Db 547 AlaGlyProProGlyPheGlnGlyLeuProGlyProSerGlyProAlaGlyGluAlaGly 566
QY 1046 CCCCCAGGTTC-----GGAAG-----CCAAGCGGGCC----- 1017
Db 567 LysProGlyGluArgGlyLeuHisGlyGluPheGlyValProGlyProAlaGlyProArg 586
QY 1016 -----GGCATG-----GACAGCAGTGGCGGCGACAGGAGGGGCGGCGGCTTCTG 969
Db 587 GlyGluArgGlyLeuProGlyGluSerGlyAlaVal-----GlyProAlaGlyProIle 604
QY 968 CTGGCTCGTGGGCGCCAGCGCTGCTCTCAGCCACCAGCAGCAGTGTGCTGCTACGCGAG 909
Db 605 GlySerArg---GlyProSerGlyProProGlyProAsp----- 616
QY 908 TCAGAAATCAGGTGTAGCAGGCGGCAAGAGAGGCTCTCTCTGGTGCCAGGT---AGG 852
Db 617 ---GlyAsnLysGlyGluProGlyAsnValGly---ProAlaGlyAlaProGlyProAla 634
QY 851 GGGCAGGCGCTGTGTCTCCAGTCAAGTGGCAGCAGGAGGTAGCCAGGAGGCGGCGGCGG 792
Db 635 GlyProGly---GlyIleProGlyGluArgGlyValAlaGlyValProGlyGly----- 651
QY 791 GACTGATCATGAAGCATAGACAGTATAGGCT-----GGCGACAGTGGT----- 747
Db 652 -----LysGlyGluLysGlyAlaProGlyLeuArgGlyAspThrGlyAlaThr 667
QY 746 -----CCGGGTGCC-----GGA 735
Db 668 GlyArgaspGlyAlaArgGlyLeuProGlyAlaIleGlyAlaProGlyProAlaGlyGly 687
QY 734 AGAGGTACAGAG-----GCAGGGCTCCAGTGGAGTGAAGC 699
Db 688 AlaGlyAspArgGlyGlyGlyGlyProAlaGlyProAlaGlyProAlaGlyAlaArgGly 707
QY 698 ACACCTGGCCACAGAAAGTCCAGCAGCCCGCCAGGATGAGCAGTGCAGCTCCAGG 639
Db 708 IleProGlyGluArgGly-----GluProGlyProValGlyProSerGlyPhe 723
QY 638 GCTTGGGATCGGCGCAGCAGCAGCCCTGCTAGCCAGCCGG----- 600
Db 724 AlaGlyProProGlyAlaAlaGlyGlnProGlyAlaLysGlyGluArgGlyProLysGly 743
QY 599 CCCTTGGGATCAGAAAGAGGCTCAGCAGGATGCCAGGACAGTGCAGGAGGAGGAGG 540
Db 744 ProLysGly---GluThrGlyProThrGly-AlaIleGlyProIleGlyAlaSerGlyPr 762
QY 539 GCGCGCGCGCCATA-----GCGTCCACGCCAGT 513
Db 762 oProGlyProValGlyAlaAlaGlyProAlaGlyProArgGlyAspAlaGlyProGly 782
QY 512 GGTCACTGGCTGAGCTAGGAGCGGGGACACAGACAGGCGCCAGCAGTGGACCAATGCCCA 453
Db 782 yMetThrGly-PhePro-GlyAlaAlaGlyArgValGlyPro-----Pro 796
QY 452 GCACCATGGTTCATGAATCTCTCTCTA---CCGCCACTTCCAGCAGCAGGCGGCGCAT 396

```



DR MIM; 267750; --  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR003129; TSPN  
 DR Pfam; PF01391; Collagen; 7.  
 DR Pfam; PF02210; TSPN; 1.  
 DR SMART; SM00282; LamG; 1.  
 DR SMART; SM00210; TSPN; 1.  
 DR Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;  
 KW Extracellular matrix;  
 KW Polymorphism.  
 FT SIGNAL 1 23  
 FT CHAIN 24 1516  
 FT CHAIN 1334 1516  
 FT DOMAIN 24 516  
 FT DOMAIN 517 550  
 FT DOMAIN 551 560  
 FT DOMAIN 561 640  
 FT DOMAIN 641 664  
 FT DOMAIN 665 786  
 FT DOMAIN 787 809  
 FT DOMAIN 810 892  
 FT DOMAIN 893 906  
 FT DOMAIN 907 948  
 FT DOMAIN 949 961  
 FT DOMAIN 962 1034  
 FT DOMAIN 1035 1044  
 FT DOMAIN 1045 1077  
 FT DOMAIN 1078 1089  
 FT DOMAIN 1090 1111  
 FT DOMAIN 1112 1118  
 FT DOMAIN 1119 1173  
 FT DOMAIN 1174 1186  
 FT DOMAIN 1187 1204  
 FT DOMAIN 1205 1516  
 FT CARBOHYD 68  
 FT CARBOHYD 129  
 FT CARBOHYD 164  
 FT CARBOHYD 691  
 FT CARBOHYD 1329  
 FT DISULFID 1366  
 FT DISULFID 1468  
 FT SITE 1095  
 FT VARSPLIC 1 180  
 FT VARSPLIC 181 215  
 FT VARIANT 1437 1437  
 FT CONFLICT 428 428  
 FT CONFLICT 841 841  
 FT CONFLICT 877 877  
 FT CONFLICT 886 886  
 FT CONFLICT 912 912  
 FT CONFLICT 933 933  
 FT CONFLICT 975 975  
 FT CONFLICT 1064 1064  
 FT CONFLICT 1084 1084  
 FT CONFLICT 1120 1120  
 FT CONFLICT 1123 1123  
 FT CONFLICT 1126 1126  
 FT CONFLICT 1206 1206  
 FT CONFLICT 1304 1304  
 FT CONFLICT 1314 1314  
 FT CONFLICT 1323 1324  
 FT CONFLICT 1443 1443  
 FT CONFLICT 1483 1483  
 SQ SEQUENCE 1516 AA; 153840 MW; 3C70F29A4476EE76 CRC64;

Alignment Scores:

Pred. No.:

Score:

Length:  
Matches:1516  
289

Percent Similarity: 29.28% Conservative: 70  
 Best Local Similarity: 23.57% Mismatches: 367  
 Query Match: 4.85% Indels: 502  
 DB: 1 Gaps: 67  
 US-09-759-143-110 (1-3410) x CALH\_HUMAN (1-1516)  
 QY 93 CAGCAGCAGGCTTTCAGCATGGCTGAGAACGTGACCGCAGCAGGCGGTGGCAGAAA 152  
 Db GlnThrThrValAlaSerLeuGlyAlaGlnThrLeuProGlySerAspSerValSerThr 501  
 QY 153 TGGCGCGCTGGCTGATTCCTAGGCAGTTGGCGGAGCAGGAGGAGAGCGCCAGCTTCT 212  
 Db 502 TrpAsp-----GlySerValArgThrProGlyGlyArgValLysGluGly 516  
 QY 213 GGAGCAGAGCCGAGCAGAGCAGTTCCTGAGCTGCCTGAACGCCGCCCTGACCCCTACCCG 272  
 Db 517 GlyLeuLysGlyGlnLysGlyGluProGlyValProGlyProGlyArgAlaGlyPro 536  
 QY 273 CCTGGC-----CCACTATGG 287  
 Db 537 ProGlySerProCysLeuProGlyLeuProCysProValSerProLeu-Gl 556  
 QY 288 T-----CCAGAGGCTGTGGGTGAGCGCCCTGC 314  
 Db 556 yProAlaGlyProAlaLeuGlnThrValProGlyProGlyProProGlyProProGl 576  
 QY 315 T-----GCGGCACCGGAAAGCCAGCTCTTGTGTGTCACCC 350  
 Db 576 yArgAspGlyThrProGlyAlaGlyGluProGlyAspProGlyGluAspGlyLysPr 596  
 QY 351 TCCTAACCTTTGGCTGAGGTGTGTTGGCGCAGGAGCATCACCTATGTGCGGCTTCGC 410  
 Db 596 oGlyAspThrGlyProGlnGlyPheProGly-----ThrPr 608  
 QY 411 TGCTGGAAGTGGGTAGAGGAAGTTTCATGACCATGGTCTGGGCAATTGGTCCAGTGC 470  
 Db 608 oGlyAspValGlyProLysGlyAspLysGlyAspProGlyValGlyGluArgGlyProPr 628  
 QY 471 TGGGCTGTGTGTGTCCTGCTCAGGCTCAGGCTGAGGCTGAGGCTGAGGCTGATG 530  
 Db 628 oGlyProGlnGlyProProGlyPro----- 636  
 QY 531 GCCCGCCCGCCCTTCAT-----CTGGGCATGTCTCT 563  
 Db 637 -ProGlyProSerPheArgHisAspLysLeuThrPheIleAspMetGluGlySerGlyPh 656  
 QY 564 TGGGCATCCTGCTGAGCCTCTTCTCATCCCAAGGCGCGGTGCTGAGGCGCTGCTGT 623  
 Db 656 eGlyGlyAspLeuGluAlaLeuArgGlyProArgGly-----ph 669  
 QY 624 GCCGGATCCAGGCCCTTGGAGCTGAGCTGCTCATCTCTGGGCGT----- 669  
 Db 669 eProGlyProProGlyProGlyValProGlyLeuProGlyGluProGlyArgPheGl 689  
 QY 670 -----GGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTC 707  
 Db 689 yValAsnSerSerAspValProGlyProAlaGlyLeu-----ProGlyValProGlyAr 707  
 QY 708 CACTGGAGGCGCTCTCTGACCTCTTCCGGGAGCC-----GG 746  
 Db 707 gGluGlyProProGlyPheProGlyLeuProGlyProProGlyProProGlyArgGluGl 727  
 QY 747 ACCACTG-----TGCCAGGCCCTACT 767  
 Db 727 yProGlyArgThrGlyGlnLysGlySerLeuGlyGluAlaGlyAlaProGly----- 745  
 QY 768 CTGTCTATGCTTCATGATCATGCTTGGGGCTGCTGGGCTGCTGCTGCTGCCATTG 827  
 Db 746 -----HisLysGlySerLysGlyAlaProGly-----ProAla----- 756  
 QY 828 ACTGGGACACCAG-----TGCCCTGGCCCTCCTA-----CCTGG 860  
 Db 828 ACTGGGACACCAG-----TGCCCTGGCCCTCCTA-----CCTGG 860



[illegible]

QY	1895	GC	TACACAGGTAGTATTGACAGAGCGACTTGGCCAAATACTACCGGTAGAAAACCTTCC	1955
Db	1049	----	-----	1049
QY	1955	AGC	CATTGGGGTGGAGGGCTCGCTCCTACCTGGGTCCAGCTCCCGCTCCT	2005
Db	1050	----	-----ProGlyPro-GlyAsnGlnGlyProProGlyProLysGlyAla	1064
QY	2006	----	-----GTTAGCCCATGGGGCTCCGGGCTGGCCGACAGTTCTGTGCTGCCAAA	2056
Db	1065	Lys	GlyGluValGlyProProGlyProGlnPhePro-PheAspPheLeu	1082
QY	2057	GTA	TGTGGCTCTCTGCTGCCACCCCTGCTGCTGAGTCCGTAGTGCACAGCTGGGG	2116
Db	1083	----	-----GlnLeuGluAlaGluMetLysGly	1090
QY	2117	CTG	GGGGTCCCTCTCTCTCTCCAGTCTTAGGGCTGCTGACTGGAGGCTTCCAA	2176
Db	1091	----	-----GluGly	1092
QY	2177	GGG	GTTTCAGTCTGGACTTATACAGGAGGCCAGAGGCTCCATGCACCTGGAATCGCG	2236
Db	1092	sgly	-----AspArgGlyAspAlaGlyGlnLysGlyGluArgG	1105
QY	2237	GG	ACTGTCAGGTGGATTACCCAGGCTCAGGGTTAACAGCTAGCCCTAGTTGAGACAC	2296
Db	1105	Yglu	ProGlyGlyGly	1110
QY	2297	AC	TAGAGAAGGTTTTTGGAGCTGAATAAACTCAGTCACCTGGTTTCCATCTCTAAG	2356
Db	1111	----	-----GlyPheGlySerSerLeuProGlyProProPro	1124
QY	2357	CCC	TAACTGCAGCTTCGTTTAATGTAGCTCTTGTCATGGGAGTTCTTAGGATGAAACA	2416
Db	1124	----	-----	1124
QY	2417	CTC	TCCATGGGATTTGAACATATGACTTATTTAGGGGAAGAGTCTCGAGGGCAACA	2476
Db	1125	Gly	ProArgGlyTyProGlyIleProGlyProLysGlyGluSer-IleArgGly	1142
QY	2477	CAC	AAGACAGGTCCCTCAGCCACAGCAGCTCTTTTTGTGATCCACCCCTCTT	2536
Db	1143	----	-----GlnProGlyProProGlyProGln	1153
QY	2537	AC	CTTTATCAGGATGTGGCTGTGTGCTCTGTGTGCCATCACAGAGACAGGCATT	2596
Db	1153	----	-----	1153
QY	2597	TAA	TATTTAACTTATTTACAAGTAGAAGGAATCCATTGCTAGCTTTTCTGTG	2656
Db	1153	----	-----	1153
QY	2657	TTG	CTGCTAATATTTCGGTAGGTGGGATCCCAACAATCAGGTCC	2706
Db	1154	--	GlyIleGlyTyGluGlyArgGlnGlyProProGlyProProGlyProGlyProp	1173
QY	2707	CTG	ATAGCTGGTCATTGGGCTGATCATTTGCCAGAATCTTCTCTCTCTGGG	2759
Db	1173	roSer	PheProGly-----ProHisArgGlnThrIleSerValProGlyProProG	1190
QY	2760	----	-----TCTGGCCCCCAAAATGCCTAACCCAGGACCTTGGAAAT	2798
Db	1190	ly	ProProGlyProProGlyProGlyThrMetGlyAlaSerSerGlyValArgLeuT	1210
QY	2799	--	TCTACTCATCCCAATGATAATTTCCAAATGCTTTACCAAGGTGTAGGTGTGAAGG	2856
Db	1210	rpAla	ThrArg-----GlnAlaMetLeuGlyGlnValHisGluValProG	1225
QY	2857	AAG	TAGAGGTGG	2870
Db	1225	lu	-----GlyTrpLeuIlePheValAlaGluGlnGluGluLeuTyValArgValGlnA	1243







Score:	309.00	Matches:	251
Percent Similarity:	30.97%	Conservative:	41
Best Local Similarity:	26.62%	Mismatches:	321
Query Match:	4.96%	Indels:	330
DB:	1	Gaps:	49

  

US-09-759-143-110 (1-3410) x CA21_BOVIN (1-1364)	
QY 2173 GAAGCCCTCCAGCAGCGCCCTAGA-----GACTGGGAGAGAGG--- 2132	
Db 23 GlnSerLeuGlnAlaThrAlaArgLysGlyProSerGlyAspArgGlyProArgGly 42	
QY 2131 ---AGAGGAGCGCCCGCCCGCCAGCTGTGCAGCTACGACCTCAGCAGCAGCGGTGGC 2075	
Db 43 GluArgGlyProGlyProPro----- 50	
QY 2074 AGCAGAGAGCCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGCCAGCCCAT 2015	
Db 50 ----- 50	
QY 2014 GGGGTAACAGGAGCGGGAGCTGGGA-----CCAGTGAGGAGGCGCCT--- 1970	
Db 51 -----GlyArgAspGlyAspGlyProGlyProGlyProGlyProPro 68	
QY 1969 ---CCACCCCATGTGCTGAAGTTTCTACGCTGAGTATTGGCCAAAG----- 1924	
Db 69 GlyProGlyLeuGlyGlyAsnPheAla-AlaGlnPheAspAlaLysGlyGlyGlyPr 88	
QY 1923 -----TCGCTCTTGTCAAACTACTACCTGTGTAGCAAGTAATGGCGACAGACC 1874	
Db 88 cGlyProMetGlyLeuMetGlyProArgGlyProArgGlyProArgGly-AlaSerGlyAlaProGlyP 108	
QY 1873 CA-----GGCCTGCGGAGACACCATATAGCAGTGCAGTGCCTGAGCTG 1826	
Db 108 toGlnGlyPheGlnGlyProGlyProGlyProGlyProGlyProGlyProAlaG 128	
QY 1825 GACATGAGCGCCATAACAGGAGTGGGGCCACCTGGGACAGCAGGA----- 1779	
Db 128 lylalaArgGlyPro-----ProGlyProProGlyLysAlaGlyLysAspGlyHisP 145	
QY 1778 -----AGGCACT 1772	
Db 145 roGlyLysProGlyArgProGlyGluArgGlyValValGlyProGlnGlyAlaArgGlyP 165	
QY 1771 ATCCAGGATGGCGAGGT---CCAGCAGATGCCCGCGCCGGAACACCTCGCTCGGT 1715	
Db 165 heProGlyThrProGlyLeuProGly-PheLysGlyLysArgGlyHisAsnGlyLeuAsp 184	
QY 1714 GGGCTCACC-----CAC 1703	
Db 185 GlyLeuLysGlyGlnProGlyAlaProGlyValLysGlyGluProGlyAlaProGlyGlu 204	
QY 1702 CACCACAGCTAGGAGACATCACAGCAGA----- 1673	
Db 205 AsnGlyThrProGlyGlnThrGlyAlaArgGlyLeuProGlyGluArgGlyArgValGly 224	
QY 1672 -----GGCCCGCAGAGCGGGGTGGAGGTGGGAGAGGCCACTGCTCCAGCAGCCAC 1619	
Db 225 AlaProGlyProAla-GlyAlaArgGlySerAspGlySerValGlyProValGlyProAl 244	
QY 1618 GTGTCCATTAGGAG-----GGAGCTCCAGC----- 1591	
Db 244 aGlyProLysSerAlaGlyProGlyPheProGlyAlaProGlyProLysGlyG 264	
QY 1590 -TTAGGGCT-----GGCAGGAAGCT 1571	
Db 264 uLeuGlyProValGlyAsnProGlyProAlaGlyProAlaGlyProArgGlyGluValG 284	
QY 1570 GTCTATCAGGTGCTCCTCAGCTACCTCCAGTGTCCCTCGGTATT----- 1521	
Db 284 yLeuProGlyLeuSerGlyProValGlyProProGlyAsnPro-GlyAlaAsnGlyLeup 304	
QY 1520 -----TGGGCAGGAACACCTGCTCTCCCGGTG 1493	

  

Db 304 roGlyAlaLysGlyAlaAlaGlyLeuProGlyValAlaGlyAlaProGlyLeuProGlyP 324	
QY 1492 GTAGAGG-----AGCCAGGTGTAGGCGCAGGATCTCAGCGCTGAGAGGT----- 1445	
Db 324 roArgGlyIleProGlyProValGlyAlaAlaGlyAlaThrGlyAlaArg-GlyLeuVal 343	
QY 1444 ---GAACCCGGTGGCGGCTGAAGCTGTCACCCAGGCCAC-----ACT 1403	
Db 344 GlyGluProGlyProAlaGlySerLysGlyGluSerGlyAsnLysGlyGluProGlyAla 363	
QY 1402 GTGGGACAGCATGTGGCAGCCGACCCACAGGGAAGTGGCCACACTGGCCAAATAGAC 1343	
Db 364 ValGlyGlnProGlyProProGlyProSerGlyGluGluGlyLysArgGly----- 380	
QY 1342 TGTCTCGAATCCCAATCGCTGCACCGCGGTCCATGACAGAGAGAACAGGAGGAGAT 1283	
Db 381 ---SerThrGlyGlu-IleGlyProAlaGlyPro-----ProGlyProP 394	
QY 1282 GGCGCACTGCAGGAACA-----GCCCGAGGTGCCCATCCGAAACCCCTTCATCATAGTG 1229	
Db 394 roGlyLeuArgGlyAsnProGlySerArgGlyLeuProGlyAlaAlaAspGlyArgAlaGlyV 414	
QY 1228 TCTCCGGGCT-----CGGTGCCCGCTCAGCTCTGGGCACGCCCTG 1187	
Db 414 aMetGlyProAlaGlySerArgGlyAlaThrGlyProAlaGlyValArgGlyProAsnG 434	
QY 1186 GTACAGCCCTCCGCCACCAAAATCCGTGTAAACAGCGTGAAGGTCTAGTGCCTCCA 1127	
Db 434 lylaspSerGlyArgProGlyGluProGlyLeuMetGlyProArgGlyPheProGlySerP 454	
QY 1126 GCTGCACAGCTCAGCCACAGAGCGCGGAGGTGCGGGCCATCGCGC----- 1077	
Db 454 roGlyAsnIleGlyProAlaGlyLysGluGlyProValGlyLeuProGlyIleAspGlyA 474	
QY 1076 -----AGCAGAGCTGGTGCAGCGCGGAAGCAGCGGCCCGCCAGGTCCGGAA 1031	
Db 474 rgProGlyProIleGlyProAlaGlyAlaArgGlyGluProGlyAsnIleGlyPheProG 494	
QY 1030 AGCAAGCGCGCGCCGATGGAGCAGTGGGCGCACAAGAGGGGGCGCCAGCCCTTC 971	
Db 494 lylProLysGlyProSerGlyAspProGlyLysAlaGlyGluLysGlyHisAlaGlyLeuA 514	
QY 970 TGTGCTGGTGGTGGCGCCAGCCCTCCCTCCTCAG-----CCAC 932	
Db 514 laglyAlaArgGlyAlaProGly-----ProAspGlyAsnAsnGlyAlaGlnGlyProP 532	
QY 931 CAGCAGTGTGGCTGCTACGCGAGTGGAGAGATGAGGTGAGCAGGCCCAAGAGGCACT- 873	
Db 532 roGlyLeuGln-----GlyValGlnGlyGlyLysGlyGluGlnGlyProA 547	
QY 872 -----CCTCCTGGGTGCCAGGTAGGGGCCAGGGCACTGGTGTCCCACTCAATGGCAGG 818	
Db 547 laglyProProGlyPheGlnGly-----LeuProGlyProAlaGlyT 561	
QY 817 CAGGAGGTAGCCCGCAGCGCCCGCCCAAGACTGATCATGAAGGCATACAGAGTAGGCGCTG 758	
Db 561 hrAlaGlyGluAlaGlyLysProGlyGlu-----ArgGlyIleProG 575	
QY 757 GCACAGTGTGT---CCGGGTCCC-----GGAAGAGTGCAGAGCAGCGCTCCAGTGG 707	
Db 575 lylGluPheGlyLeuProGlyProAlaGlyAlaArgGlyGluArg---GlyProProGlyG 594	
QY 706 AGTGAAGACACCTGGCCACAGAAAGTCCAGACGCCCGCCAGGATGAGCAGTGCAGG 647	
Db 594 LuSerGlyAlaAlaGlyProThrGlyProIleGlySerArgGlyProSerGlyProProG 614	
QY 646 CTCCA-----GGGCGCTGGGATCCGGGCACAGCAG 617	
Db 614 lylProAspGlyAsnLysGlyGluProGlyValValGlyAla-----ProGlyThrAlaG 632	
QY 616 CCCTGTGTAGCAGCGCGCCCTTGGGATGAGAAAGAGCTCAGCAGGATGCCAAGGACAG 557	









KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KF Glycoprotein; Collagen; Signal.

FT SIGNAL 1 24 POTENTIAL.

FT PROPEP 25 ? AMINO-TERMINAL PROPEPTIDE (POTENTIAL).

FT CHAIN 2 1096 COLLAGEN ALPHA 2(I) CHAIN.

FT PROPEP 1097 1356 CARBOXYL-TERMINAL PROPEPTIDE

FT CARBOHYD 1257 1257 (BY SIMILARITY).

FT SEQUENCE 1356 AA; 126985 MW; 7BB2F1F80DB10C93 CRC64;

Alignment Scores:

Pred. No.: 3,31e-07 Length: 1356

Score: 308.00 Matches: 235

Percent Similarity: 32.39% Conservative: 51

Best Local Similarity: 26.61% Mismatches: 319

Query Match: 4.9% Indels: 279

DB: 1 Gaps: 46

US-09-759-143-110 (1-3410) x CA21\_ONCMY (1-1356)

QY 2101 CTACGCCACCTCAGCAGCAGCGGTGGCAGAGAGCCACATTACTTTGGCACAACAGA 2042

DB 18 LeuLeuAlaSerGlnSerGlyGlyLeuLysGlyProArg-----GlyAlaLys 34

QY 2041 AACTGGCGGCCAGCCCGCAGCCCATGGGCTAAC---AGGAGCGGGAGCTGGGA--- 1988

DB 35 GlyProArgGlyAspArgGlyProGlnGlyProAsnGlyArgAspGlyLysAlaGlyLeu 54

QY 1987 CCCAGTGAGGAGCCCTC-----CCACCCCAATGTCTGGAAATTTCT----- 1944

DB 55 ProGlyIleAlaGlyProGlyProGlyLeuGlyGlyAsnGlyAlaAlaGlnPhe 74

QY 1943 -----ACGCTGATGATTTGGCCAACTCGCTCT 1917

DB 75 AspGlyGlyLysGlySerAspProGlyProMetGlyLeuMetGlySerArgGly 94

QY 1916 TGTCAATACTACTCTGTAGCAAGTAATGGCAGCACCA-----GGC 1869

DB 95 ProAsnGlyProPro-----GlyAlaProGlyProGlnGlyPheThrGly 109

QY 1868 CTGGCGGACACCATATAGCAGTACAGACTGGCTGAGCTGAGCAATGGAGCCCATAA 1809

DB 110 HisAlaGlyGluProGlyGluProGlyGlnThrGlySerIleGlyAlaArgGlyProThr 129

QY 1808 ACAGGATGGGCCACCTGGGACAGCAGGAGGAC-----TATCAGGATGG 1761

DB 130 GlySerAlaGlyLysProGlyGluAspGlyAsnGlyArgProGlyLysProGlyAsp 149

QY 1760 CGA---GGTCCAGGACAGATGCCCGGCCCGGAAACCCAC-----CCTGGCTCGGTGG 1713

DB 150 ArgGlyGlyProGly-ThrGlnGlyAlaArgGlyPheProGlyThrProGlyLeuProGly 169

QY 1712 GCTACCCACACACAGTACGAGGACATCAGAGCAG----- 1673

DB 169 yMetLysGlyHisArgGlyTyAsnGlyLeuAspGlyArgLysGlyGluSerGlyThrAl 189

QY 1673 ----- 1673

DB 189 aGlyAlaLysGlyGluThrGlyAlaHisGlyAlaAsnGlySerProGlyProAlaGlyse 209

QY 1672 -----GGCCCCGAGCGCGGTG 1653

DB 209 rArgGlyLeuAsnGlyGluArgGlyArgAlaGlyProAlaGlyProAla-GlyAlaArg 229

QY 1652 GAGGTGGGAGCAGCCACTGCTCCAGCACCACCGCTGTCTTATAGG----- 1606

DB 229 lyAlaAspGlySerThrGlyProAlaGlyProAlaGlyProLeuGlyAlaAlaGlyProp 249

QY 1605 -----AAGCGAGCTCCAGGC----- 1591

DB 249 roGlyPheProGlyAlaProGlyProLysGlyGluIleGlyAlaGlySerAsnGlyp 269

QY 1590 -----TTAG 1587

DB 269 roSerGlyProGlnGlyArgGlyGluProGlyIleAsnGlyAlaValGlyProValG 289

QY 1586 GGCCT-----GGCAGGAAGCTGGTCTCAGCTAGTGT 1557

DB 289 lyProValGlyAsnProGlyAsnAsnGlyIleAsnGlyAlaLys-GlyAlaAlaGlyLeu 308

QY 1556 CCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTTGGCAGGAAACACCTGCTTCC 1497

DB 309 Pro-----GlyVal---AlaGlyAlaProGlyPhePro 318

QY 1496 GGTGTAGAGGAGCCAGCTGTGTAGG-----GCAGGATCTGCAGGCGTG 1452

DB 319 GlyProArgGlyGlyProGlyProGlnGlySerThrGlyAlaArgGlyLeu 338

QY 1451 AGAAGGTGAACCC-----GGTGAGGCGGCTGAAGCTGTCAACACGG 1410

DB 339 -GlyGlyAspProGlyProSerGlyGlnLysGlyAspSerGlyAlaLysGlyGluProGly 358

QY 1409 CCACACT-----GTGGACAGGCAATGGCCAGCCAGCCAGCCAGCTG 1356

DB 358 yHisSerGlyValGlnGlyAlaAlaGlyProAlaGlyGluGlyLysArgGlySerThr 378

QY 1355 TGGCCAAATAGACTCTCGAGTGGCGAATCGCTGCACCCAGCGGCTCATGACAGAGAGA 1296

DB 378 rGlyGluValGlyAlaThrGlyProAlaGlyLeu-----ArgGly 391

QY 1295 AGACAGGAGGATGGCGCACTGCAGGAACAGCCAGGCTG-----CCCA 1251

DB 391 yAlaArgGlyGly-AlaGlyThrArgGlyLeuProGlyLeuGlyGlyArgGlyProI 411

QY 1250 TCCGAACGCTTCATCATAGTCTCCGGCTCGCTCGGCTCGCTCGGCTCAGCTCTGGCAGCG 1191

DB 411 leGlyMetProGlyAla-----ArgGlyAlaThrGlyProGlyGlyIleArgGlyAla- 428

QY 1190 CTGTGTACAGCCCTCGCCACAGAAATCGGTAAACAGCGCTGAAGCTCATGAGTGCCA 1131

DB 429 ProGlyAspAlaGlyArgAlaGlyGluSerGlyLeuThrGlyAlaArgGlyLeuProGly 448

QY 1130 TCCAGCTGCAGCTCAGCCAGAGAGCCGCGGAGGTGCGGGCATGCGGC----- 1077

DB 449 AsnSerGlyGlnGlyProGlyLysGlyGluProGlyAlaAlaGlyLeuAsp 468

QY 1076 -----AGCACAGCTGTGTGAGCCGGGAGAGCGCGCCAGGTTC----- 1035

DB 469 GlyArgThrGlyProGlyProGlyProThrGlyProArgGlyGlnProGlyAsnIleGlyPhe 488

QY 1034 GGAAGCCAAAGCGGCGCATGAGCAGTGGGCGGACAGAGGGGGCGGCGCAGACGCC 975

DB 489 ProGlyProLysGlyProGlyGlyGluAlaGlyLysGlyGlyAspGlyProThrGly 508

QY 974 CTTCCTGCTCGCTGGTGGGCGCCAGCGCTG-----CCTCCTCAGCCACCA 930

DB 509 AlaThrGlyLeuArgGlyGlyProGlyAlaAspGlyAsnAsnGlyAlaProGlyProAla 528

QY 929 GCAGTGTGCTGTACAGAGTGAAGATGAGGTGAGCAGGCAAGAGGCACTCTCT 870

DB 529 GlyValVal-----GlyAsnThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 542

QY 869 CTGGTGTGCTCGCTG-----AGGGGCGCAGGCGCTGCTGCTCCAGTCAATGG 822

DB 543 AlaGlyAlaProGlyPheGlnGlyLeuProGlyProAlaGlyProAlaGlyGluAlaGly 562

QY 821 CAGGAGGAGGTAGTCCAGCGAGCCCAAGACTGATCATGAGCATAGACAGAGTAGG 762

DB 563 LysAlaGly-----AsnGlnGlyMet 569

QY 761 CTGGCGCAGCTGGT---CCGGGTCCCGGA-----AGAGGTGACAGAGA----- 723

DB 570 ProGlyAspGlnGlyLeuProGlyProAlaGlyValLysGlyGlyGlyGlyGlyGlyGly 589

QY 722 ---GCAGGCGCTCCAGTGGAGTGAAGCACACCTGCGCCAGAGTCCAGCAGCCCGCC 666

Db 590 ProAlaGlySerGlnGlyAlaIleGlyAlaArgGlyProAlaGly---Thr 608  
QY 665 CCAGGATGAGCAGTGCAGCTCCAGGCGCTGGGATCCGGCCACAGCAGCCCTGTAGCC 606  
Db 609 ProGly-----ProAspGlyGlyLysGlyGluProGlySerValGlyIleValGly 625  
QY 605 AGCCGG-----CCCTTGGGA-----TGAGAAAGAGGCTCAGCAGATGCCCA 564  
Db 626 AlaAlaGlyHisGlnGlyProGlyGlyMetProGlyGluArgGlyAlaGlyGlyThrPro 645  
QY 563 AGCAGAGTCCAGATGAAGCGCGCGCGCCATAGGCTCCAGCCAGTGGTCACTGG 504  
Db 646 Gly-----ProLysGlyGluLysGlyGluGlyHisArgGlyLeuGluGlyAsnMet 663  
QY 503 CTGAGCTAGGAGCGGACACAGACAGCCAGCCAGCTGGACCAATGCCCAGCACCATTGG 444  
Db 664 GlyArgAspGlyAlaArgGlyGlyProGlyPro----- 674  
QY 443 TCATGAACCTCTCTCTACCCCACTT----- 417  
Db 675 -----SerGlyProProGlyProSerGlyAlaAsnGlyGluLysGlyGluSerGlySer 692  
QY 416 -----CCAGCAGCAGCGGCACATAGTGCATCGCTCGCGCCAAACACACCTCCAGCC 363  
Db 693 PheGlyProAlaGlyProAlaGlyLeuArg-----GlyProSerGly 706  
QY 362 CAAAGCTTAGCAGGTTGACCAAGAGCTGGGCTTTCGGTCCGCGCCAGCGCGCTCA 303  
Db 707 GluArgGlyGluGlyPro-----AlaGlyLeuProGlyPheAla-----GlyPro 722  
QY 302 CCCACAGCTCTGGACCATAGTGGCGGCGGAGGCTCAGGCGCGCTCAGGAC 243  
Db 723 ProGlySerAspGly-GlnSerGlyProArgGly--GluLysGlyProAlaGlyGlyLy 741  
QY 242 TCCAGAACCTCTCGTCTCGCTCGCTCCAGAGCTCGCGCTCTCTCTCTCTCTCTCT 183  
Db 741 sGlyAspValGlyProAlaGlyProAlaGlyProSerGlyGlnSerGlyProSerGlyAl 761  
QY 182 CCAACTGCTTAGGAATCAGCAGGCGCGCCATTTCTGCCAGCCCTTGTGGCGGCTCCAGC 123  
Db 761 aSerGlyProAlaGlyProProGly-----GlyArgGlyAspAl 774  
QY 122 TTCTACGCCATGCTCAACA-----CCTGCTGTGTGGGCGCACCTCAGTGGGACAGCT 69  
Db 774 aGlyProSerGlyLeu-ThrGlyPheProGlyAlaAlaGlyArgValGlyGly----- 791  
QY 68 CTCATCACTCAGATCTGCGCGCGCGCGCGCTGTCACCGCGGACCGCGCG 16  
Db 792 -----ProGlyProAlaGlyIleAlaGlyProProGlySerAla 804  
RESULT 32  
CA2B\_MOUSE  
ID CA2B\_MOUSE STANDARD; PRT; 1650 AA.  
AC O64739; 0921W0;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Collagen alpha 2(XI) chain precursor.  
GN COL11A2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SvJ;  
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,  
RA Hall J., Lasky S., Hood L.;  
RT \*Sequence of the mouse major histocompatibility locus class II  
RT region.\*;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-1592 FROM N.A.

RC STRAIN=FVB/N, and 129/Sv; TISSUE=Cartilage;  
RX MEDLINE=97135795; PubMed=8981332;  
RA Vandenberg P., Vuorio M.M., Ala-Kokko L., Prockop D.J.;  
RT "The mouse colla2 gene. Some transcripts from the adjacent rxr-beta  
gene extend into the colla2 gene."  
RL Matrix Biol. 15:359-367(1996).  
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS (BY SIMILARITY).  
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(XI) IS A POST-TRANSLATIONAL  
CC OF ALPHA 3(XI)=1(XI) (BY SIMILARITY).  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY  
CC ALTERNATIVE SPLICING.  
CC -!- PM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or  
CC send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AF100956; AAC69905.1; -  
CC EMBL; U16789; AAA67751.1; -  
CC EMBL; U16790; AAA67752.1; -  
CC MGD; MGI:88447; Coll1a2.  
CC InterPro; IPR000087; Collagen.  
CC InterPro; IPR000885; Fib\_collagen\_C.  
CC InterPro; IPR001791; Laminin\_G.  
CC InterPro; IPR001230; Prenyl\_site.  
CC InterPro; IPR003129; TSPN.  
CC Pfam; PF01391; Collagen; 18.  
CC Pfam; PF01410; COLFI; 1.  
CC Pfam; PF02210; TSPN; 1.  
CC ProDom; PD000007; Collagen; 4.  
CC ProDom; PD002078; Fib\_collagen\_C; 1.  
CC SMART; SM00038; COLFI; 1.  
CC SMART; SM00282; LamG; 1.  
CC SMART; SM00210; TSPN; 1.  
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
CC Glycoprotein; Collagen; Alternative splicing; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 1414  
FT PROPEP 1415 1650  
FT DOMAIN 23 255  
FT DOMAIN 487 1414  
FT DOMAIN 1415 1650  
FT CONFLICT 450 450  
FT CONFLICT 618 619  
FT CONFLICT 711 711  
FT CONFLICT 757 759  
FT CONFLICT 768 768  
FT CONFLICT 790 790  
FT CONFLICT 803 803  
FT CONFLICT 836 836  
FT CONFLICT 919 919  
FT CONFLICT 1167 1167  
FT CONFLICT 1300 1300  
FT CONFLICT 1436 1436  
SQ SEQUENCE 1650 AA; 162072 MW; 8ECDB8702E71E2DA CRC64;  
COLLAGEN ALPHA 2(XI) CHAIN.  
CARBOXYL-TERMINAL PROPEPTIDE.  
NONHELICAL REGION.  
TRIPLE-HELICAL REGION.  
NONHELICAL REGION.  
R -> L (IN REF. 2).  
NQ -> KP (IN REF. 2).  
V -> A (IN REF. 2).  
TGP -> HGS (IN REF. 2).  
A -> S (IN REF. 2).  
R -> G (IN REF. 2).  
G -> V (IN REF. 2).  
G -> D (IN REF. 2).  
E -> V (IN REF. 2).  
P -> S (IN REF. 2).  
A -> T (IN REF. 2).  
I -> M (IN REF. 2).  
ALIGNMENT SCORES:  
Pred. No.: 3.59e-07  
Score: 307.00  
Length: 1650  
Matches: 315  
Percent Similarity: 31.06%  
Conservative: 59  
Best Local Similarity: 26.16%  
Mismatch: 379  
Query Match: 4.93%  
Indels: 454  
Gaps: 73  
DB:





RT "Identification of a distinct type IV collagen alpha chain with  
RT restricted kidney distribution and assignment of its gene to the  
RT locus of X chromosome-linked Alport syndrome.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
RN [6]  
RN SEQUENCE OF 1442-1471 FROM N.A.  
RX MEDLINE=90252791; PubMed=2339699;  
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D.,  
RA Sheer D., Solomon E., Pihlajaniemi T.;  
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
RT to the region of the X chromosome containing the Alport syndrome  
RT locus.";  
RL Am. J. Hum. Genet. 46:1024-1033(1990).  
RN [7]  
RN SEQUENCE OF 1-20 FROM N.A.  
RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,  
RA Marynen P.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RN SEQUENCE OF 1258-1270 FROM N.A. (SPliced FORM).  
RX MEDLINE=94133540; PubMed=8301933;  
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,  
RA Cassiman J.-J., Marynen P.;  
RT "Differential splicing of COL4A5 mRNA in kidney and white blood  
RT cells: a complex mutation in the COL4A5 gene of an Alport patient  
RT deletes the NCI domain.";  
RL Kidney Int. 44:1316-1321(1993).  
RN [9]  
RN REVIEW ON VARIANTS.  
RX MEDLINE=97338662; PubMed=9195222;  
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
RT "The clinical spectrum of type IV collagen mutations.";  
RL Hum. Mutat. 9:477-499(1997).  
RN [10]  
RN VARIANT AS SER-1564.  
RX MEDLINE=91169492; PubMed=1672282;  
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,  
RA Tryggvason K.;  
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a  
RT conserved cysteine to serine in Alport syndrome.";  
RL Genomics 9:10-18(1991).  
RN [11]  
RN VARIANT AS ARG-325.  
RX MEDLINE=92303559; PubMed=1376965;  
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,  
RA Tryggvason K., Gubler M.-C., Antignac C.;  
RT "Substitution of arginine for glycine 325 in the collagen alpha 5  
RT (IV) chain associated with X-linked Alport syndrome: characterization  
RT of the mutation by direct sequencing of PCR-amplified lymphoblast  
RT CDNA fragments.";  
RL Am. J. Hum. Genet. 51:135-142(1992).  
RN [12]  
RN VARIANT AS GLU-325.  
RX MEDLINE=93244772; PubMed=1363780;  
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,  
RA Rizzoni G.F., de Marchi M.;  
RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
RT glutamic acid in Alport syndrome.";  
RL Hum. Mol. Genet. 1:127-129(1992).  
RN [13]  
RN VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
RX MEDLINE=94010948; PubMed=8406498;  
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,  
RA Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
RA van Oost B.A., Smeets H.J.M.;  
RT "Identification of four novel mutations in the COL4A5 gene of  
RT patients with Alport syndrome.";  
RL Genomics 17:485-489(1993).  
RN [14]  
RN VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872  
RX AND C-1241.  
RA MEDLINE=95322976; PubMed=7599631;  
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;

RT "Detection of 12 novel mutations in the collagenous domain of the  
RT COL4A5 gene in Alport syndrome patients.";  
RL Hum. Mutat. 5:197-204(1995).  
RN [15]  
RN VARIANT AS ARG-1649.  
RX MEDLINE=96213750; PubMed=8651292;  
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
RA Denison J.C., Fain P.R., Gregory M.C.;  
RT "A mutation causing Alport syndrome with tardive hearing loss is  
RT common in the western United States.";  
RL Am. J. Hum. Genet. 58:1157-1165(1996).  
RN [16]  
RN VARIANTS AS.  
RX MEDLINE=96213754; PubMed=8651296;  
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,  
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
RA Savi M., Ballabio A., de Marchi M.;  
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
RT exons of the COL4A5 gene.";  
RL Am. J. Hum. Genet. 58:1192-1204(1996).  
RN [17]  
RN VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
RX MET-1428.  
RX MEDLINE=97094179; PubMed=8940267;  
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jäcasier D.,  
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
RA Gubler M.-C., Antignac C.;  
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
RT syndrome.";  
RL Am. J. Hum. Genet. 59:1221-1232(1996).  
RN [18]  
RN VARIANT AS ASP-1498.  
RX MEDLINE=96233932; PubMed=8829632;  
RA Tverskaya S., Bobrynya V., Tsalykova F., Ignatova M.,  
RA Krasnopol'skaya X., Evgrafov O.;  
RT "Substitution of Al498D in noncollagen domain of alpha 5(IV) collagen  
RT chain associated with adult-onset X-linked Alport syndrome.";  
RL Hum. Mutat. 7:149-150(1996).  
RN [19]  
RN VARIANT AS GLN-1677.  
RX MEDLINE=97295089; PubMed=9150741;  
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
RT "Common ancestry of three Ashkenazi-American families with Alport  
RT syndrome and COL4A5 R1677Q.";  
RL Hum. Genet. 99:681-684(1997).  
RN [20]  
RN VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.  
RX MEDLINE=98112435; PubMed=9452056;  
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
RA Trivelli A., Pinciaroli A.R., Ragalolo M., Rizzoni G.F., de Marchi M.;  
RT "Missense mutations in the COL4A5 gene in patients with X-linked  
RT Alport syndrome.";  
RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
RN [21]  
RN VARIANTS AS V-420; 456-P-P-458 DEL; D-573; D-624; D-635; 802-G--P-807  
RX DEL; R-869; C-941; S-1030; S-1066; D-1143; R-1196; E-1261; S-1357  
AND R-1649.  
RX MEDLINE=99063529; PubMed=9848783;  
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,  
RA Barker D.F., Gregory M.C., Atkin C.L., Strykarsdottir U., Neumann H.,  
RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;  
RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
RT Alport syndrome using PCR and direct DNA sequencing.";  
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
RN [22]  
RN VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1158;  
RX SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
RX MEDLINE=20030197; PubMed=10561141;  
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,  
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;  
RT "Detection of mutations in the COL4A5 gene in over 90% of male







RA Apte S., Mattei M.-G., Olsen B.R.;  
 RT "Cloning of human alpha 1(X) collagen DNA and localization of the  
 RL COL10A1 gene to the q21-q22 region of human chromosome 6.";  
 RN FEBS Lett. 282:393-396(1991).  
 [7]  
 RA SEQUENCE OF 547-655 FROM N.A.  
 RP MEDLINE-92077285; PubMed-1743401;  
 RX Reichenberger E., Aigner T., von der Mark K., Stoeck H., Bertling W.;  
 RA "In situ hybridization studies on the expression of type X collagen  
 RT in fetal human cartilage.";  
 RL Dev. Biol. 148:562-572(1991).  
 RN [8]  
 RA REVIEW ON VARIANTS.  
 RP MEDLINE-97255959; PubMed-9101290;  
 RX Kuivaniemi H., Tromp G., Prockop D.J.;  
 RA "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RL cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RN Hum. Mutat. 9:300-315(1997).  
 [9]  
 RA VARIANTS SMCD ASP-598 AND PRO-614.  
 RP MEDLINE-94136476; PubMed-8304336;  
 RX Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,  
 RA Grant M.E., Boot-Handford R.P.;  
 RT "Amino acid substitutions of conserved residues in the  
 RL carboxyl-terminal domain of the alpha 1(X) chain of type X collagen  
 RN occur in two unrelated families with metaphyseal chondrodysplasia  
 type Schmid.";  
 [10]  
 RA Am. J. Hum. Genet. 54:169-178(1994).  
 [11]  
 RA VARIANTS SMCD ARG-591.  
 RP MEDLINE-94272470; PubMed-8004099;  
 RX McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;  
 RA "Additional mutations of type X collagen confirm COL10A1 as the  
 RT Schmid metaphyseal chondrodysplasia locus.";  
 RL Hum. Mol. Genet. 3:303-307(1994).  
 [12]  
 RA VARIANTS SMCD VAL-618.  
 RP MEDLINE-95181449; PubMed-7876225;  
 RX Chan D., Cole W.G., Rogers J.G., Bateman J.F.;  
 RA "Type X collagen multimer assembly in vitro is prevented by a Gly618  
 RT to Val mutation in the alpha 1(X) NCI domain resulting in Schmid  
 RL metaphyseal chondrodysplasia.";  
 [13]  
 RA J. Biol. Chem. 270:4558-4562(1995).  
 [14]  
 RA VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.  
 RP MEDLINE-95331767; PubMed-7607655;  
 RX Bonaventure J., Chaminade F., Maroteaux P.;  
 RA "Mutations in three subdomains of the carboxy-terminal region of  
 RT collagen type X account for most of the Schmid metaphyseal  
 RL dysplasias.";  
 [15]  
 RA Hum. Genet. 96:58-64(1995).  
 [16]  
 RA VARIANTS SMCD PRO-600.  
 RP MEDLINE-96375754; PubMed-8782043;  
 RX Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,  
 RA Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;  
 RT "Mutations within the gene encoding the alpha 1(X) chain of type X  
 RL collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but  
 RN not several other forms of metaphyseal chondrodysplasia.";  
 [17]  
 RA J. Med. Genet. 33:450-457(1996).  
 [18]  
 RA VARIANTS SMCD GLU-18 AND ARG-18.  
 RP MEDLINE-97220591; PubMed-9067753;  
 RX Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;  
 RA "Mutations in the N-terminal globular domain of the type X collagen  
 RT gene (COL10A1) in patients with Schmid metaphyseal  
 RL chondrodysplasia.";  
 [19]  
 RA Hum. Mutat. 9:131-135(1997).  
 [20]  
 RA VARIANTS SMD GLU-595.  
 RP MEDLINE-99057503; PubMed-9837818;  
 RX Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,

RA Nakamura Y.;  
 RT "Mutation of the type X collagen gene 'COL10A1' causes  
 RL spondylometaphyseal dysplasia.";  
 RN Am. J. Hum. Genet. 63:1659-1662(1998).  
 [16]  
 RA VARIANT SMCD CYS-597.  
 RP MEDLINE-99069781; PubMed-9852679;  
 RX Sawai H., Ida A., Nakata Y., Koyama K.;  
 RA "Novel missense mutation resulting in the substitution of tyrosine by  
 RT cysteine at codon 597 of the type X collagen gene associated with  
 RL Schmid metaphyseal chondrodysplasia.";  
 RN J. Hum. Genet. 43:259-261(1998).  
 [17]  
 RA FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC  
 RP CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
 RX MINERALIZATION ZONES OF HYALINE CARTILAGE.  
 [18]  
 RA SUBUNIT: HOMOTRIMER.  
 RP PM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 RX UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 [19]  
 RA DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE  
 RX METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED  
 RP DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE  
 RX PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.  
 RP RADIOGRAPHY USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE  
 RX METAPHYSES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE  
 RP KNEES.  
 [20]  
 RA DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDYLOMETAPHYSEAL  
 RP DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE  
 RX SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE  
 RP VERTEBRAL BODIES OF THE SPINE AND METAPHYSES OF THE TUBULAR BONES.  
 [21]  
 RA SIMILARITY: STRONG. TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
 [22]  
 RA SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
 [23]  
 RA THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 RP between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 RX the European Bioinformatics Institute. There are no restrictions on its  
 RP use by non-profit institutions as long as its content is in no way  
 RX modified and this statement is not removed. Usage by and for commercial  
 RP entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 RX or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 [24]  
 RA EMBL; X60382; CAA42933.1; -  
 RP EMBL; X65120; CAA46236.1; -  
 RX EMBL; X98568; CAA67178.1; -  
 RP EMBL; AL121963; CAB87590.1; -  
 RX EMBL; X68531; RAC60615.1; -  
 RP EMBL; X58879; CAA41686.1; -  
 RX EMBL; M74050; AAA61221.1; -  
 RP EMBL; X72579; CAA51170.1; -  
 RX EMBL; X72580; CAA51170.1; JOINED.  
 RP PIR; S15826; S15826.  
 RX PIR; S30086; S30086.  
 RP PIR; A43901; A43901.  
 RX PIR; S18249; S18249.  
 RP PIR; S21856; S21856.  
 RX PIR; S26396; S26396.  
 RP Genew; HGNC:2185; COL10A1.  
 RX MIM; 120110; -  
 RP MIM; 156500; -  
 RX MIM; 184250; -  
 RP InterPro; IPR001073; C1q.  
 RX InterPro; IPR000087; Collagen.  
 RP Pfam; PF00386; C1q; 1.  
 RX Pfam; PF01391; Collagen; 8.  
 RP PRINTS; PRO0007; COMPLEMENTC1Q.  
 RX SMART; SM00110; C1q; 1.  
 RP PROSITE; PS01113; C1q; 1.  
 RX Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 RP Cartilage; Collagen; Signal; Disease mutation; Polymorphism.  
 [25]  
 RA SIGNAL 1 18 POTENTIAL.  
 RP CHAIN 19 680 COLLAGEN ALPHA 1(X) CHAIN.  
 RX DOMAIN 19 56 NONHELICAL REGION (NC2).  
 RP DOMAIN 57 519 TRIPLE-HELICAL REGION.  
 RX DOMAIN 520 680 NONHELICAL REGION (NC1).



Db 453 -ProIleGly---ProProGlyIleProGlyPhePro----- 463  
 QY 540 CGGGGGCCATAGCTCCACCCAGTGGTCACTAGGCTAGGAGCGGGACACAG 481  
 Db 464 -----GlySerLysGlyAs 468  
 QY 480 ACCAGGCCAGCACTGGACCAATGCCAGCAGCATGGTCATGAACCTCTCTACCC 421  
 Db 468 pProGly-----SerProGlyProPr 475  
 QY 420 ACTTCCAGCAGAGCGGCACATAGGTGATCGCTCGCGCCAAACACACCTCCAGGCCA 361  
 Db 475 oGlyProAlaGly---IleAlaThrLysGlyLeuAsnGlyProThrGlyProGlyPr 494  
 QY 360 AAGGTTAGCAGGTTGACCAAGAGAGCTGGCTTTCGGTCCGCGAGCAGGCGGCTCACC 301  
 Db 494 o-----ProGlyProArgGlyHisSer-----GlyGluPr 504  
 QY 300 CACAGCCTTGGACCATAGTGGCCAGCGGGTCTAGGGCTCAGGGGGCCGCTTCAGGCACCTC 241  
 Db 504 oGlyLeuProGlyPro-ProGlyProPro-----GlyProGlyGlnAlaValMetP 522  
 QY 240 CAGAA---CTGCTTCTCGCTCGCTCTCCAGAACCTCGCGCCCTCTCTCTCTGCTGCC 184  
 Db 522 roGluGlyPheIleLysAlaGlnArgProSerLeuSerGlyThrPro-LeuValSer 541  
 QY 183 GCCAAGCTAGTGGTATCAGCCAGCGCCGCTTCTGCGAGCCCTTGTGTCGGCTCCAG 124  
 Db 542 AlaAsn---GlnGlyValThrGlyMetProValSerAlaPheThrValIleLeuSerLys 560  
 QY 123 CTCCTCAGCCCATGCTCAGCTGCTGTGTGGGACCTCAGTGGGAGACACGCTCAT 64  
 Db 561 AlaTyrProAlaIleGlyThrProIleProPheAspLysIleLeuTyrAsnArgGlnGln 580  
 QY 63 CACTCAGATCCTGGCGGAGGC 43  
 Db 581 HisTyrAspProArgThrGly 587  
 RESULT 35  
 CA1A\_HUMAN STANDARD; PRT; 680 AA.  
 AC Q03692;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 GN Collagen alpha 1(X) chain precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92109659; PubMed=1764025;  
 RA Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,  
 RA Solomon E., Grant M.E., Boot-Handford R.P.;  
 RT "The human collagen X gene. Complete primary translated sequence and  
 RT chromosomal localization.";  
 RL Biochem. J. 280:617-623(1991).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93012005; PubMed=1397333;  
 RA Reichenberger E., Beier F., Luvall P., Olsen B.R., von der Mark K.,  
 RA Bertling W.M.;  
 RT "Genomic organization and full-length cDNA sequence of human collagen  
 RT X.";  
 RL FEBS Lett. 311:305-310(1992).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Beier F., Lamm M.B., von der Mark K.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.

RA Williams S.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE OF 52-680 FROM N.A.  
 RX MEDLINE=92267014; PubMed=1587271;  
 RA Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;  
 RT "Cloning of the human and mouse type X collagen genes and mapping of  
 RT the mouse type X collagen gene to chromosome 10.";  
 RL Eur. J. Biochem. 206:217-224(1992).  
 [6]  
 RP SEQUENCE OF 561-666 FROM N.A.  
 RX MEDLINE=91243838; PubMed=2037056;  
 RA Apte S., Mattei M.-G., Olsen B.R.;  
 RT "Cloning of human alpha 1(X) collagen DNA and localization of the  
 RT COL10A1 gene to the q21-q22 region of human chromosome 6.";  
 RL FEBS Lett. 282:393-396(1991).  
 [7]  
 RP SEQUENCE OF 547-655 FROM N.A.  
 RX MEDLINE=92077285; PubMed=1743401;  
 RA Reichenberger E., Aigner T., von der Mark K., Stoeb H., Bertling W.;  
 RT "In situ hybridization studies on the expression of type X collagen  
 RT in fetal human cartilage.";  
 RL Dev. Biol. 148:562-572(1991).  
 [8]  
 RP REVIEW ON VARIANTS  
 RX MEDLINE=97255959; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 [9]  
 RP VARIANTS SMCD ASP-598 AND PRO-614.  
 RX MEDLINE=94136476; PubMed=8304336;  
 RA Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,  
 RA Grant M.E., Boot-Handford R.P.;  
 RT "Amino acid substitutions of conserved residues in the  
 RT carboxyl-terminal domain of the alpha 1(X) chain of type X collagen  
 RT occur in two unrelated families with metaphyseal chondrodysplasia  
 RT type Schmid.";  
 RL Am. J. Hum. Genet. 54:169-178(1994).  
 [10]  
 RP VARIANT SMCD ARG-591.  
 RX MEDLINE=94272470; PubMed=8004099;  
 RA McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;  
 RT "Additional mutations of type X collagen confirm COL10A1 as the  
 RT Schmid metaphyseal chondrodysplasia locus.";  
 RL Hum. Mol. Genet. 3:303-307(1994).  
 [11]  
 RP VARIANT SMCD VAL-618.  
 RX MEDLINE=95181449; PubMed=7876225;  
 RA Chan D., Cole W.G., Rogers J.G., Bateman J.F.;  
 RT "Type X collagen multimer assembly in vitro is prevented by a Gly618  
 RT to Val mutation in the alpha 1(X) NC1 domain resulting in Schmid  
 RT metaphyseal chondrodysplasia.";  
 RL J. Biol. Chem. 270:4558-4562(1995).  
 [12]  
 RP VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.  
 RX MEDLINE=95331767; PubMed=7607655;  
 RA Bonaventure J., Chaminade F., Maroteaux P.;  
 RT "Mutations in three subdomains of the carboxy-terminal region of  
 RT collagen type X account for most of the Schmid metaphyseal  
 RT dysplasias.";  
 RL Hum. Genet. 96:58-64(1995).  
 [13]  
 RP VARIANT SMCD PRO-600.  
 RX MEDLINE=96375754; PubMed=8782043;  
 RA Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,  
 RA Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;  
 RT "Mutations within the gene encoding the alpha 1(X) chain of type X  
 RT collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but  
 RT not several other forms of metaphyseal chondrodysplasia.";  
 RL J. Med. Genet. 33:450-457(1996).

[14]  
 RP VARIANTS SMCD GLU-18 AND ARG-18.  
 RA MEDLINE=97220591; PubMed=9067753;  
 RX Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;  
 RT "Mutations in the N-terminal globular domain of the type X collagen  
 gene (COL10A1) in patients with Schmid metaphyseal  
 chondrodysplasia.";  
 RL Hum. Mutat. 9:131-135(1997).  
 [15]  
 RP VARIANTS SMD GLU-595.  
 RX MEDLINE=99057503; PubMed=9837818;  
 RA Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,  
 RA Nakamura Y.;  
 RT "Mutation of the type X collagen gene 'COL10A1' causes  
 spondylometaphyseal dysplasia.";  
 RL Am. J. Hum. Genet. 63:1659-1662(1998).  
 [16]  
 RP VARIANT SMCD CYS-597.  
 RX MEDLINE=99069781; PubMed=9852679;  
 RA Sawai H., Ida A., Nakata Y., Koyama K.;  
 RT "Novel missense mutation resulting in the substitution of tyrosine by  
 cysteine at codon 597 of the type X collagen gene associated with  
 Schmid metaphyseal chondrodysplasia.";  
 RL J. Hum. Genet. 43:259-261(1998).  
 CC -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC  
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.  
 CC -!- SUBUNIT: HOMOTRIMER.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE  
 CC METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED  
 CC DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE  
 CC PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND WADDLING GAIT.  
 CC RADIOGRAPHY USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE  
 CC METAPHYSES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE  
 CC KNEES.  
 CC -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDYLOMETAPHYSEAL  
 CC DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE  
 CC SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE  
 CC VERTEBRAL BODIES OF THE SPINE AND METAPHYSES OF THE TUBULAR BONES.  
 CC -!- SIMILARITY: STRONG. TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
 CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X60382; CAA42933.1; -  
 CC EMBL; X65120; CAA46236.1; -  
 CC EMBL; X98568; CAA67178.1; -  
 CC EMBL; AL121963; CAB87590.1; -  
 CC EMBL; S68531; AAC60615.1; -  
 CC EMBL; X58879; CAA41686.1; -  
 CC EMBL; M74050; AAG61221.1; -  
 CC EMBL; X72579; CAA51170.1; -  
 CC EMBL; X72580; CAA51170.1; JOINED.  
 CC PIR; S15826; S15826  
 CC PIR; S30086; S30086.  
 CC PIR; A43901; A43901.  
 CC PIR; S18249; S18249.  
 CC PIR; S21856; S21856.  
 CC PIR; S26396; S26396.  
 CC Genew; HGNC:2185; COL10A1.  
 CC MIM; 120110; -  
 CC MIM; 156500; -  
 CC MIM; 184250; -  
 CC InterPro; IPR001073; Clq.  
 CC InterPro; IPR000087; Collagen.

DR Pfam; PF00386; Clq; 1.  
 DR Pfam; PF01391; Collagen; 8.  
 DR PRINTS; PRO0007; COMPLENTC1Q.  
 DR SMART; SM00110; Clq; 1.  
 DR PROSITE; PS01113; Clq; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cartilage; Collagen; Signal; Disease mutation; Polymorphism.  
 FT SIGNAL 1 18  
 FT CHAIN 19 680  
 FT DOMAIN 19 56  
 FT DOMAIN 57 519  
 FT DOMAIN 520 680  
 FT DOMAIN 545 680  
 FT VARIANT 18 18  
 FT VARIANT 18 18  
 FT VARIANT 545 545  
 FT VARIANT 591 591  
 FT VARIANT 591 591  
 FT VARIANT 591 591  
 Alignment Scores:  
 Pred No.: 5,09e-07 Length: 680  
 Score: 304.50 Matches: 184  
 Percent Similarity: 34.57% Conservative: 39  
 Best Local Similarity: 28.53% Mismatches: 232  
 Query Match: 4.74% Indels: 190  
 DB: 1 Gaps: 38  
 US-09-759-143-110 (1-3410) x CALA\_HUMAN (1-680)  
 QY 129 CCGCACCAAGGCTGGCAGAAATGGCGCCCTGGCTGATTCCTAGCAGCTTGGCGCAG 188  
 DB 74 ProGlyProSerGly---ProProGlyLysProGly-----TyrGlySerProGlyLeu 90  
 QY 189 CAAGGAGGAGGCGCCGAGCTTCTGGAGCAGAGCCAGACGA---AGCAGTCTTGGAGTG 245  
 DB 91 GlnGlyLysProGlyLeuProGlyProProGlyProSerAlaValGlyLysProGlyVal 110  
 QY 246 CCTGAAGCGCCCTGAGCCCTACCCGCTGGCCCTATGCTGCTGCTGCTGCTGCTGCTGCT 297  
 DB 111 ProGlyLeuProGlyLysProGlyGluArgGlyPro-TyrGlyProLysGlyAspValG 130  
 QY 298 -----GTGGTGAGCGCGCTCTCGCTGCTGGA-----AGTGGGGTAGAGGAGA 326  
 DB 130 yProAlaGlyLeuProGlyProAArgGlyProProGlyProProGlyProGlyProAl 150  
 QY 327 AAGCCAGCTCTGCTGCTCAACCTGCTAACCTTTGGCTGGAGGCTGTGTTGGCCCGCAG 386  
 DB 150 aGlyLeuSerValProGlyLysProGlyGlnGlnGlyProThrGlyAlaProGlyProAr 170  
 QY 387 GCATCACCCTATGTCCCGCTCTGCTGCTGGA-----AGTGGGGTAGAGGAGA 434  
 DB 170 gGlyPheProGlyGluLysGlyAlaProGlyValProGlyMetAsnGlyGlnLysGlyG 190  
 QY 435 AGTTCATGACCATGCTGGGCGATGCTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 488  
 DB 190 uMet-----GlyTyrGlyAlaProGlyArgProGlyGluArgGlyLeuPr 205  
 QY 489 CGCTCTAGGCTCAGCCAGTACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548  
 DB 205 oGlyProGlnGlyProThrGlyProSerGly-----ProProGlyValG 220  
 QY 549 TCTGGGCACTGCTTGGGCGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 605  
 DB 220 yLysArgGlyGluAsnGlyValProGlyGlnProGlyLysGlyAspArgGlyPhePr 240  
 QY 606 GCGTAGCAGGCTGCTGTG-----CCCGGATGCCAGCCCTGGAGCTGGCAGCTGCTGCTGCT 659  
 DB 240 oGlyGluMetGlyProLysGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 254  
 QY 660 TCCTGGG---CGTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 716

```

Db 255 -ProGlyGluAArgGlyProGluGlyLeuGlyProGlyAlaAlaGlyAlaProGlyG1 274
QY 717 CCTGCTCTCTGACCT-----CTTCCGGGA-----CC 743
Db 274 nProGlyIle-----ProGlyThrLysGlyLeuProGlyAlaProGlyIleAlaGlyPro 293
QY 744 CGGACCACTG-----TGCCAGGCTACTCTGTCTATGCTTCCATGATCATGCTG 794
Db 293 oGlyProProGlyPheGlyLysProGlyLeuProGlyLeuLysGlyGluAArgGlyProAl 313
QY 795 GGGGCTGCTGGCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 854
Db 313 aGlyLeuProGlyGlyProGlyAla-----LysGlyGluGlnGlyProAlaGlyLe 330
QY 855 ACTTGGGACCAAGCA-----GGAGTGCCTCTTTGGGCTGCTCACCTCATCTTCCTCA 908
Db 330 uProGlyLysProGlyLeuThrGlyProPro----- 340
QY 909 CCTGCTAGCAGCACACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 968
Db 341 -----GlyAsnMetGlyProGln----- 346
QY 969 CAGAGGCTGTGCGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1028
Db 347 -----GlyPro-----LysGlyIleProGlySerHisG1 356
QY 1029 TTTTCCGGAAGCTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1088
Db 356 yLeuProGlyProLysGlyGluThrGlyProAlaGlyProAlaGlyTyPro----- 373
QY 1089 GCACCTCGCGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1148
Db 374 ----GlyAlaLysGlyGluAArgly----- 380
QY 1149 CGCTGTTTACAGGATTTCTGTCGGAGGGCTGTACAGGGCTGCTGCTGCTGCTGCTGCTG 1208
Db 381 -----SerProGlySerAspGlyLysProGlyTyProGlyLys-----Pr 394
QY 1209 CGGGACCGAGCGCGGACACTATGATGATGATGATGATGATGATGATGATGATGATGATG 1263
Db 394 oGlyLeuAspGlyPro-----LysGlyAsnProGlyLeuProGlyProLy 409
QY 1264 -GCTGTTCTCAGTGGCCCATCTCCCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1322
Db 409 sGlyAspProGlyValGlyProGlyProGlyLeuProGlyProValGlyProAlaGlyAl 429
QY 1323 AGCGATTGCGGCACTCGAGCAGCTATTTGGCAGCTGTGCGAGCTTTCCTGCTGCTGCTG 1382
Db 429 a-----LysGlyMetProGlyHisAsnGlyGlu-----AlaGlyProAr 442
QY 1383 GTGCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1439
Db 442 gGlyAlaProGlyIleProGlyThr---ArgGly--ProIleGlyProProGlyIlePro 460
QY 1440 GTTTCACCTTCTCAGCCTCGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1499
Db 461 GlyPheProGly--SerLysGlyAspPro-----GlySerProGlyProGly 476
QY 1500 AGAAGCAGGTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1559
Db 476 yProAlaGlyIleAlaThrLysGlyLeuAsnGly----- 487
QY 1560 CCCTGATGACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1619
Db 488 -----ProThr--GlyPro---ProGlyProProGlyProArgGlyHis 501
QY 1620 TGGGTGTGGAGGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1679
Db 501 erGlyGluProGly-----LeuProGlyProPro----- 510
QY 1680 GTGATGTCTCGTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1739

```

```

Db 511 -----GlyProProGlyPro-----ProGlyG 518
QY 1740 GGGCATCTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1799
Db 518 ln-----AlaValMetProGlyGlyPheIleLysAlaGlyGlnArgP 532
QY 1800 CATCCCTGTTATGGGCTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1859
Db 532 roSerLeuSerGlyThrProLeuValSerAlaAsnGlnGlyValThrGlyMetProVal 552
QY 1860 CTGCCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913
Db 552 erAlaPheThrValIleLeuSerLysAlaTyProAlaIleGlyThrProIleProPhe 572
QY 1914 ACAAG 1918
Db 572 sPlys 573

RESULT 36
CA24_HUMAN
ID CA24_HUMAN STANDARD; PRT; 1712 AA.
AC P08572;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=89086769; PubMed=3198637;
RA Hostikka S.L., Tryggvason K.;
RT "The complete primary structure of the alpha 2 chain of human type IV
RT collagen and comparison with the alpha 1(IV) chain.";
RL J. Biol. Chem. 263:19488-19493(1988).
RN [2]
RP SEQUENCE OF 1-1042 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88151998; PubMed=3345760;
RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;
RT "Human basement membrane collagen (type IV). The amino acid sequence
RT of the alpha 2(IV) chain and its comparison with the alpha 1(IV)
RT chain reveals deletions in the alpha 1(IV) chain.";
RL Eur. J. Biochem. 172:35-42(1988).
RN [3]
RP SEQUENCE OF 1254-1712 FROM N.A.
RX MEDLINE=87219158; PubMed=3582677;
RA Hostikka S.L., Kurkinen M., Tryggvason K.;
RT "Nucleotide sequence coding for the human type IV collagen alpha 2
RT chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
RT (IV) but not with the collagenous domain or 3'-untranslated region.";
RL FEBS Lett. 216:281-286(1987).
RN [4]
RP SEQUENCE OF 1451-1485 FROM N.A.
RX MEDLINE=87092438; PubMed=3025878;
RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
RT alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
RN [5]
RP SEQUENCE OF 1486-1712 FROM N.A.
RX MEDLINE=87250571; PubMed=2439508;
RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
RT "Duplication of type IV collagen COOH-terminal repeats and species-
RT specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
RL J. Biol. Chem. 262:9231-9238(1987).
RN [6]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;

```

"The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region.";

J. Biol. Chem. 263:17217-17220(1988).

[7]

SEQUENCE OF 1-33 FROM N.A.

RP MEDLINE=89030632; PubMed=2846280;

RA Poeschl E., Pollner R., Kuehn K.;

"The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure.";

EMBO J. 7:2687-2695(1988).

[8]

SEQUENCE OF 1-33 FROM N.A.

RP TISSUE-Skin;

RC MEDLINE=93305049; PubMed=8317999;

RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;

"Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription.";

Biochem. J. 292:687-695(1993).

[9]

SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.

TISSUE=Placenta;

RC MEDLINE=89005112; PubMed=2844531;

RA Siebold B., Deutzmann R., Kuehn K.;

"The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal, non-collagenous aggregation and cross-linking domain of basement-membrane type IV collagen.";

Eur. J. Biochem. 176:617-624(1988).

[10]

"FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/NIDOGEN.

-1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -

ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE

WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

-1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS

DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE

G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY

CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL

TRIPLE-HELICAL 7S DOMAIN.

-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING

UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

-1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH

ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF

THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE

IV COLLAGENS.

[11]

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[12]

EMBL; X05562; CAA29076.1;

EMBL; X05610; CAA29098.1;

EMBL; J02760; AAA58422.1;

EMBL; X36963; AAA53099.1;

EMBL; X12784; CAA31275.1;

EMBL; J04217; AAA53097.1;

PIR; A32024; A32024.

Genew; HGNC:2203; COL4A2.

MIM; 120090;

InterPro; IPR000087; Collagen.

InterPro; IPR001442; Procollagnc4.

Pfam; PF01391; Collagen; 24.

Pfam; PF01413; C4; 2.

ProDom; PD000007; Collagen; 5.

ProDom; PD003923; Procollagnc4; 2.

SMART; SM00111; C4; 2.

Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KW Glycoprotein; Basement membrane; Collagen; Signal.

[13]

FT SIGNAL 1 25

FT PROPEP 26 183

FT CHAIN 184 1712

FT DOMAIN 184 1484

FT DOMAIN 1485 1712

FT DISULFID 1504 1593

FT DISULFID 1537 1590

FT DISULFID 1549 1555

FT DISULFID 1612 1708

FT DISULFID 1646 1705

FT DISULFID 1658 1665

FT CARBOHYD 138 138

FT CONFLICT 471 471

FT CONFLICT 683 683

FT CONFLICT 1575 1575

FT CONFLICT 1663 1663

FT CONFLICT 1701 1701

FT CONFLICT 1701 1701

SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Alignment Scores:

pred. No.: 4 86e-07

Score: 304.00

Percent Similarity: 28.95%

Best Local Similarity: 23.63%

Query Match: 4.74%

Indels: 522

Gaps: 70

DB: 1

US-09-759-143-110 (1-3410) x CA2A\_HUMAN (1-1712)

QY 95 GCAGCAGGTGTTGAGCATGGCTGAGACCTGGACCGGCACCAAGGGCTGCAGANAATG 154

Db 95 AlaProGlyValThr---GlyProGlyAspValGlyAlaArgValSerGlyPhe 113

QY 155 ---GGCGCCTGGCTGATCTCTAGGCAGTTGGCGGAGCAAGGA-----GGAGAGGCCGC 205

Db 114 ProGlyAlaAspGlyLeuPro-GlyHisProGlyGlnGlyGlyProArgGlyArgProGly 133

QY 206 AGCTTCTGGAGCAGACCGCAGAGAGAGAGTCTGGAGTGGCTGACGACGCCGCCCT----- 260

Db 133 yTyAspGlyCysAsnGlyThrGlnGlyAspSerGly---ProGlnGlyProGlyGlySe 152

QY 261 -GAGCCCTAG-----CCGCTGGCCCACTATGCTCCAGAGCTGTGGTGGAGCCGCCCTG 313

Db 152 rGluGlyPheThrGlyProGlyProGlyPro-GlnGlyProGlyGlnGlyGlyGluProT 172

QY 314 CTGCGGCACCG-----GAAAGCCCGAGCTCTTGGTCAACCTGCTAACCTTTGGCCCTG 367

Db 172 yrAlaLeuProGlyGluGluArgAspArgTyrArgGlyGluProGlyGlu-----ProG 190

QY 368 GA-----GGTGTGTTGGCGCAGGCATCACCTATGTCGCCGCTCTGCTG 412

Db 190 lyLeuValGlyPheGlnGlyProGlyArgProGlyHisValGlyGlnMetGlyProV 210

QY 413 CTGGAAGTGGCGGTAGA-----GGAGAGTTCATG 442

Db 210 alGlyAlaProGlyArgProGlyProGlyProGlyProGlyGlnGlnGlyA 230

QY 443 ACCATGCTGCTGGGCATTTGG-----TCCAGTGTCTGGGC 475

Db 230 snArgGlyLeuGlyPhePheTyrGlyValGlyGlyGlyGlyGlyGlyGlnProGlyP 250

QY 476 CTGCTGCTGCTCCGCTCTCTAGCTAGCTAGCCAG-----TGACCCATGGCGTGCACGCTAT 529

Db 250 roAsnGlyIleProSerAspThrLeuHisProIleIleAlaProThrGlyValThrPheH 270

QY 530 GGCGCCCG----- 537

Db 270 isProAspGlnTyrLysGlyGluLysGlySerGluGlyGluProGlyIleArgGlyIles 290

QY 538 -----CGGCCCTTCAT-----CTGG 553





QY	2225	ACTGGAATCGGGGACTCTGCAGGTGCGATTACCCAGGCTCAGGTTTAACACAGCTAGCCTCC	2284
Db	908	yle-----AspGlyMetProGlyThrProGlyLeuLysGlySerGlySerP	925
QY	2285	TAGTTTGAGACACACCTAGAGAAGGGTTTTGGGAGCTGAATAAACTCAGCTCA---CCTGG	2341
Db	925	oGlyMetAsp-----GlyPheGlnGlyMetProGlyLeuLysGlyArgProG	941
QY	2342	TTTTCCCATCTCTAAGCCCTTAACCTGCAGCTCGHTTAATGTAGCTCTTGCATGGGAGT	2401
Db	941	yPheProGly-----	944
QY	2402	TTCTAGGATGAACACTCTCTCCATGGCATTTGAACATATGACTTATTTTAGGGGAAGAG	2461
Db	945	-----SerLysGlyGlnAlaGlyPhePheGlyIleProGlyLeu-LysGlyLeuA	961
QY	2462	TC-----CTGAGGGCAACACACAGAACACCGTCCCTCAGCCACACGACT	2509
Db	961	laGlyGluProGlyPheLysGlySerArgGlyAspProGlyPro-----	975
QY	2510	GTCTTTTGTGTGATCATCCGCCCTCTTACCTTTTATCAGGATGTGCGCTTGGTCTCTTC	2569
Db	976	-----ProGlyProProPro-----ValI	982
QY	2570	TGTTGCCATCACAGACACAGCGCATTTAAATATTAACTTATTATTAACAAAGTAGA	2629
Db	982	leLeuProGlyMetLysAsp-----	988
QY	2630	AGGGAATCCATCTGCTAGCTTTCTGTGTTGGTGTCTAATATTGGGTAGGGTGGGGGATC	2689
Db	989	-----lleLysGlyGluLysGlyAsp-----	995
QY	2690	CCCAACAATCAGTCCCTCGAGATAGCTGGTCAATGGGCTGATCATTCGCAGAACTTCT	2749
Db	996	-----GluGlyProMet-----	999
QY	2750	TCTCTGGGTCTGGCCCCCAAAATGCCTAACCCAGGACCTTGGAAATTTCTACTCATCC	2809
Db	1000	-----GlyLeu-----	1001
QY	2810	CAATGATAATTCCAATGCTGTGTACCCAGGTTAGGTGTTCAAGGAAGGTAGAGGGTG	2869
Db	1002	-----LysGlyTyrLeuG-----	1006
QY	2870	GGGCTTCAGTCTCAACGGCTTCCCTAACCAACCCCTCTTCTCTTGGCCAGCGCTGGTTC	2929
Db	1006	lyAlaLysGlyIleGlnGlyMetProGlyIlePro-----GlyLeuSerGlyIleP	1023
QY	2930	CCCCACTTCACCTCCCTCTACTCTCTCTAGGACTGGGCTGATGAGGCACTGCCCAAA	2989
Db	1023	roGlyLeuPro-----GlyArgProGlyHisIleLysGlyValLysGlyA	1038
QY	2990	TTTCCCCTACCCCAACTTTTCCCTACCCCAACTTTCCCAACCAAGCTCCCAACCCCTGT	3049
Db	1038	spIleGlyValProGlyIleProGlyLeuProGlyPhePro-----	1051
QY	3050	TTGAGGACTACTCGAGGACAGACAAAGTCGGGTTTCCCAAGCCTTTGTTCATCTCA	3109
Db	1052	-----GlyValAlaGlyProProGlyIleThr--GlyPheProGlyPheIleGlySerAr	1069
QY	3110	GCCCCCAGAGTATATCTGCTGTGGGGAATCTCACACAGAAATTCAGAGCACCCCTCG	3169
Db	1069	g-----GlyAspLysGlyAlaPro-----	1075
QY	3170	CTGAGCTAAGGAGGTCTTATCTCTCAGGGGGG	3203
Db	1076	----GlyValAlaGlyLeuTyrGlyGluIleGly	1085

37 JUL 26 1964

## RESULT 37

### CALB HUMAN

CA1B_HUMAN	STANDARD:	PRT: 1806 AA.
TD CA1B HUMAN		

ID	CALB_HUMAN	STANDARD;	PRT;	180
BC	R12107: Q14034:	Q9UIT4:	Q9UIT5:	Q9UIT6:

AC FI2107; Q14034; Q90114; Q90115  
DT 01-OCT-1989 (Rel. 12; Created)







CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

DR EMBL; 222964; CA80536.1; -  
 DR EMBL; 222964; CA80537.1; -  
 DR EMBL; J05066; AA27989.1; -  
 DR EMBL; J05066; AA27989.1; -  
 DR EMBL; U22327; AAA64312.1; ALT\_SEQ.  
 DR EMBL; U53342; AAA96215.1; -  
 DR EMBL; U53342; AAA96216.1; -  
 DR PIR; A34476; A34476.  
 DR WormPep; F01G12.5a; CE04334.  
 DR WormPep; F01G12.5b; CE04335.  
 DR InterPro; IPR001442; ProcollagenC4.  
 DR Pfam; PF01391; Collagen; 23.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD000007; Collagen; 1.  
 DR ProDom; PD003923; ProcollagenC4; 2.  
 DR SMART; SM00111; C4; 2.  
 DR Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;  
 KW Alternative splicing; Glycoprotein; Signal.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1758  
 FT DOMAIN 27 42  
 FT DOMAIN 42 1527  
 FT DOMAIN 1528 1758  
 FT DISULFID 1546 1635  
 FT DISULFID 1591 1597  
 FT DISULFID 1654 1750  
 FT DISULFID 1688 1747  
 FT DISULFID 1700 1707  
 FT CARBOHYD 248 248  
 FT VARSPPLIC 229 264

FT VARIANT 48 48  
 FT VARIANT 366 366  
 FT VARIANT 570 570  
 FT VARIANT 588 588  
 FT VARIANT 597 597  
 FT VARIANT 690 690  
 FT VARIANT 690 690  
 FT VARIANT 737 737  
 FT VARIANT 877 877  
 FT VARIANT 904 904  
 FT VARIANT 1003 1003  
 FT VARIANT 1125 1125  
 FT VARIANT 1152 1152  
 FT VARIANT 1286 1286  
 FT CONFLICT 1604 1604  
 FT CONFLICT 1682 1682  
 FT SEQUENCE 1758 AA; 167750 MW; 97EE3F3DBB2D2AC5 CRC64;

Alignment Scores:  
 Pred. No.: 5,1e-07  
 Score: 303.50  
 Percent Similarity: 31.55%  
 Best Local Similarity: 24.96%  
 Query Match: 4.73%  
 DB: 1

Length: 1758  
 Matches: 318  
 Conservative: 84  
 Mismatches: 435  
 Indels: 438  
 Gaps: 73

US-09-759-143-110 (1-34110) x CA24\_CABEL (1-1758)

QY 26 CCGGTGTACAGCCGCGCGCTCGGCACGAGTCTGAGTGTAGTGTAGTGTGTCCCTGAGG 85  
 Db |||||: ||| |||  
 QY 409 ProGlyAsnGluGlyLeuProGlyProLysGly-AspGlyAspGlyGly 425  
 QY 86 TGGCCACAGCAGCAGCTGTGAG- - - - -CATGGCTGAGAAGCTGGACGGCACCACAA 139  
 Db ||| ||| ||||| |||||: ||| ||| |||||  
 QY 426 IleProGlyAlaProGlyValSerGlyProSerGlyIleProGlyLeu-ProGlyProly 445  
 QY 140 GGC- - - - -TGCAGAAATGGCGCTGCTGATTCTCTAGGACGTTGGCGCAGCAAGG 193  
 Db |||||: ||||| ||||| |||||: |||||  
 QY 445 sGlyGluProGlyTyrArgGlyThrProGlyGlnSerIleProGlyLeuProGlyLysAs 465  
 QY 194 AGGAGAGCGCGAGCTTCTGGAGCAGCGGAGACGAGCAAGCAGTCTTGGAGTGCCTGAGG 253  
 Db |||||: ||||| ||||| |||||: |||||: |||||  
 QY 465 pGlyLysProGlyLeuAspGlyAlaProGlyArgLysGlyGluAsnGlyLeu- - - - - 482  
 QY 254 GCGCCCTGAGCCCTACCGCTGCGCCACTATGTGTCAGAGGCTGTGGGTGAGCGCCCTG 313  
 Db ||||| ||||| ||||| ||||| |||||: |||||  
 QY 483 - - - - -ProGlyVal-ArgGlyPro- - - - -GlyAspSerLeuA 494  
 QY 314 CTGCGCAGCGGAAAGCCAGCTCTTGTGCTGTCACCT- - - - - 351  
 Db ||| ||||| ||||| ||||| |||||  
 QY 494 sNglyLeuProGlyAlaProGlyGlnArgGlyAlaProGlyProAsnGlyTyrAspGlyA 514  
 QY 352 - - - - -GCTAACCTTGGCCTGGA- - - - -GGTGTGTTGGCCGACGATCCTAT 397  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 514 rGAspGlyValAsnGlyLeuProGlyAlaProGlyThrLysGlyAspArgGlyGlyThrC 534  
 QY 398 GTGCGCGCTCTGCTGTGGAAGTGGGGTGA- - - - - 429  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 534 ySerAlaCysAlaProGlyThrLysGlyGluLysGlyLeuProGlyTyrSerGlyGlnP 554  
 QY 430 - - - - - - - - - - -GGAGAAGTTCAG 442  
 QY 554 roGlyProGlnGlyAspArgGlyLeuProGlyMetProGlyProValGlyAspAlaGlyA 574  
 QY 443 ACCATGCT- - - - -GCTGGCATTGGTCCAGCTGCTGGCCCT- - - - -G 478  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 574 sPAspGlyLeuProGlyProAlaGlyArgProGlySerProGlyProGlyGlnAspG 594  
 QY 479 GTCTGTCTCCGCTCTAGGCTCAGCCAGTACACCTGGCGTGGCGTGGCGTGGCGCCGCG 538  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 594 lPheProGlyLeuProGlyGlnLysGlyGluProThrGlnLeuThrLeuArgProGlyP 614  
 QY 539 CGGCGCTTCATCTGGGCACT- - - - -GTCTTGGGCATCTCTGTGAGCCCT- - - - - 582  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 614 roProGlyTyrProGlyLeuLysGlyGluAsnGlyPheProGlyGlnProGlyValAspG 634  
 QY 583 - - - - -CTTTCTCATCCCAAGCGCGCTGCT- - - - - 609  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 634 lYLeuProGlyProSerGlyProValGlyProGlyAlaProGlyTyrProGlyGluL 654  
 QY 610 - - - - -ACCAGGCTGTGCTGCGCGGA- - - - -TCCAGCGCGCTGGA- - - - - 645  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 654 yGlyAspAlaGlyLeuProGlyLeuSerGlyLysProGlyGlnAspGlyLeuProGlyL 674  
 QY 646 - - - - -GCTGCACCTGCTCATCTCTGGG- - - - - 666  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 674 euProGlyAsnLysGlyGluAlaGlyTyrGlyGlnProGlyGlnProGlyPheProGlyA 694  
 QY 667 - - - - -CTGGGGCTGTGGACTTCTGTGGCAGGTGTGCTTCACTCCACTGGAGGCGCTGCTC 724  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 694 lALysGlyAspGlyGlyLeu- - - - -ProGlyLeu- - - - -ProGlyThrProGlyL 709  
 QY 725 TCTGACCTCTTCCGGGACCC- - - - -GGACCACTGTGCGCAGGCC 763  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 709 euGlnGlyMetProGlyGluProAlaProGluAsnGlnValAsnProAlaProGly- 728  
 QY 764 TACTCTGTCTATGCTTCATGATCAGTCTTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCC 823  
 Db ||||| ||||| ||||| ||||| ||||| |||||  
 QY 729 - - - - -GlnProGlyLeuProGlyLeuPro- - - - - 736  
 QY 824 ATTGACTGGGACACCAAGTGCCTGCGCCCTACCTGCGGCCACCCAGGA- - - - - 870







QY 2776 GCCTAACCCAGGACCTTGGAAATCTACTCATCCCAATGATTAATTCCTTCTTAC 2835  
 Db 1332 ysArgGlyGluAsp---Gly-LeuProGlyValProGlyArgAspGlyGlnProGlyIle 1350  
 QY 2836 CCAAGGTAGGCTTCAAGGAAGTAGAGGGTGGGCTTCAGGTCTCTC----- 2883  
 Db 1351 ProGlyLeu-----LysGlyGluValGlyGlyGlyGlyGlyGlyGlnPro 1366  
 QY 2884 -----AACGGCTTCCTAACCCACCCCTCTTCT 2910  
 Db 1367 GlyPheProGlyIleProGlyLeuLysGlyGluGlyGlyGlyGlyGlyPheProGlyAla 1386  
 QY 2911 CTTGGCCAGCCTGGTTCCTCCCACTTCCA----- 2940  
 Db 1387 LysGlyGluAlaGlyPheProGlyThrProGlyValProGlyTyrAlaGlyGluLysGly 1406  
 QY 2941 -----CPCCTCTCTACTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAAT 2991  
 Db 1407 AspGlyGlyLeuProGlyLeuProGlyArgAspGlyLeu----- 1419  
 QY 2992 TCCCTACCCCACTTTCCTTCCCTACCCCACTTTCCTCCCACTCCCAACCCCTGTT 3051  
 Db 1420 ---ProGlyAlaAspGlyProValGlyProGlyProSer----- 1432  
 QY 3052 GGAGCTACTGAGGACCAAGCAAGCAAGTGCCTTCCCAAGCCTTTGTCCTATCTCAGC 3111  
 Db 1433 -----Gly 1433  
 QY 3112 CCCAGAGATATCTGTCTGGGGAATCTCACACAGAACTCAGGACCCCTCCCT 3171  
 Db 1434 ProGlnAsnLeu-----ValGluProGlyGlyLysGlyLeuPro 1446  
 QY 3172 GAGCTAAGGAGGCTTATCTCTCAGGGG 3201  
 Db 1447 GlyLeuProGlyAlaProGlyLeuArgGly 1456

RESULT 39  
 CA21.RAT  
 ID CA21.RAT STANDARD; PRT; 1372 AA.  
 AC P02456; Q9R1E8;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(I) chain precursor.  
 GN COL1A2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Guenther D., Seibold S., Marx M.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 86-98.  
 RC TISSUE-Skin;  
 RX MEDLINE=67162268; PubMed=5337886;  
 RA Kang A.H., Bornstein P., Piez K.A.;  
 RT "The amino acid sequence of peptides from the cross-linking region of  
 rat skin collagen";  
 RL Biochemistry 6:788-795(1967).  
 RN [3]  
 RP SEQUENCE OF 99-102.  
 RC TISSUE-Skin;  
 RX MEDLINE=69206881; PubMed=5785232;  
 RA Fietzek P.P., Piez K.A.;  
 RT "Isolation and characterization of the cyanogen bromide peptides from  
 the alpha 2 chain of rat skin collagen";  
 RL Biochemistry 8:2129-2133(1969).  
 RN [4]  
 RP SEQUENCE OF 102-144.  
 RC TISSUE-Skin;

RX MEDLINE=73049496; PubMed=4636752;  
 RA Fietzek P.P., Kell I., Kuehn K.;  
 RT "The covalent structure of collagen. Amino acid sequence of the N-  
 terminal region of alpha 2-CB4 from calf and rat skin collagen.";  
 RL FEBS Lett. 26:66-68(1972).  
 RN [5]  
 RP SEQUENCE OF 423-452.  
 RC TISSUE-Skin;  
 RX MEDLINE=71115216; PubMed=5344653;  
 RA Higberger J.H., Kang A.H., Gross J.;  
 RT "Comparative studies on the amino acid sequence of the alpha 2-CB2  
 peptides from chick and rat skin collagens";  
 RL Biochemistry 10:610-616(1971).  
 RN [6]  
 RP SEQUENCE OF 453-501.  
 RC TISSUE-Skin;  
 RX MEDLINE=75059250; PubMed=4435743;  
 RA Fietzek P.P., Kuehn K.;  
 RT "The covalent structure of collagen: amino acid sequence of the N-  
 terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5  
 from calf skin collagen";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).  
 RN [7]  
 RP SEQUENCE OF 791-836.  
 RC TISSUE-Skin;  
 RX MEDLINE=74055004; PubMed=4763308;  
 RA Fietzek P.P., Kuehn K.;  
 RT "The covalent structure of collagen: amino acid sequence of the N-  
 terminal region of alpha 2-CB5 from rat skin collagen";  
 RL FEBS Lett. 36:289-291(1973).  
 RN [8]  
 RP ORDER OF CNBR PEPTIDES.  
 RX MEDLINE=70181852; PubMed=5443712;  
 RA Vuust J., Lane J.M., Fietzek P.P., Miller E.J., Piez K.A.;  
 RT "The order of the CNBR peptides from the alpha 2 chain of collagen";  
 RL Biochem. Biophys. Res. Commun. 38:703-708(1970).  
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 (FIBRILLAR FORMING COLLAGEN).  
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 HYDROXYAPATITE.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF121217; AA041775.1;  
 DR PIR; A02867; CGRT2S.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; Fib-collagen\_C.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Pfam; PF01410; COLFI; 1.  
 DR ProDom; PD000007; Collagen; 3.  
 DR ProDom; PD002078; Fib-collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 24  
 FT PROPEP 25 85  
 FT Amino-TERMINAL PROPEPTIDE  
 FT (BY SIMILARITY).  
 FT CHAIN 86 1108  
 FT COLLAGEN ALPHA 2(I) CHAIN  
 FT PROPEP 1109 1372  
 FT CARBOXYL-TERMINAL PROPEPTIDE  
 FT (BY SIMILARITY).  
 FT SITE 783 785  
 FT CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 828 830  
 FT CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 1011 1013  
 FT CELL ATTACHMENT SITE (POTENTIAL).  
 FT

FT	MOD_RES	86	86	PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT	MOD_RES	90	90	CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT				INVOLVED IN CROSS-LINKING.
FT	CARBOHYD	1273	1273	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	132	132	T -> P (IN REF. 4).
FT	CONFLICT	137	137	S -> P (IN REF. 4).
FT	CONFLICT	145	422	MISSING (IN REF. 4).
FT	CONFLICT	431	432	ST -> TS (IN REF. 5).
FT	CONFLICT	494	494	E -> Z (IN REF. 6).
FT	CONFLICT	497	497	N -> A (IN REF. 6).
FT	CONFLICT	502	790	MISSING (IN REF. 6).
FT	CONFLICT	825	825	R -> K (IN REF. 7).
SQ	SEQUENCE	1372 AA;	129564 MW;	B069371A8DB20A72 CRC64;

  

Alignment Scores:			
Pred. No.:	6.le-07	Length:	1372
Score:	302.00	Matches:	302
Percent Similarity:	31.04%	Conservative:	55
Best Local Similarity:	26.26%	Mismatches:	393
Query Match:	4.85%	Indels:	402
DB:	1	Gaps:	61

  

US-09-759-143-110 (1-3410) x CA2L_RAT (1-1372)			
QY	3074	GCTTCTGTCCTGCAGTAGCTCCAAACAGGGTTGTGGAGCTGTGGGGAAAGTTGGGGT	3015
DB	233	AlaProGlyProAlaGlyAlaArg	245
QY	3014	AGGGAAAGTTGGGGTAGGGAAATTTTGGCGAGTGC	2976
DB	246	ValGlyProValGlyProAlaGlyProIleGlySerAlaGlyProGlyPheProGly	265
QY	2975	-----TTTCATCAGCCAGTCCTAGAGAGATAGAGGGAGTGGAAGT	2934
DB	266	AlaProGlyProLysGlyGluLeuGlyProVal	280
QY	2933	GGGGGAACAGGCTGGCCACAGAGAGAGGGTGTGGTGGAGAGCGTTGGACCTGAA	2874
DB	281	ProAlaGlyProAlaGlyProArgGlyGluAlaGlyLeuProGlyLeuSerGlyProVal	300
QY	2873	GCCTCCACCTTACCTTCTTCAACACCCCTAACTTGGGTACAGCATTTGGTAATTCATCA	2814
DB	301	GlyProProGly-----AsnProGlyAlaAsnGlyLeuThrGlyAla	314
QY	2813	TTTGGGATGAGTAGAATTTCAAGGTCCTGGGTAGGCATTTTGGGGGGCCACGCCAG	2754
DB	315	LysGlyAlaThrGlyLeuProGlyValAlaGly---Ala-ProGlyLeuProGlyProAr	333
QY	2753	GAGAAGAAGATTCTGGCAATGATCAGCCCAATGACCATCTACTCAGGGGACCTG	2699
DB	333	g-----GlyIleProGlyProValGlyAl	341
QY	2698	-----ATTGTGGGATCCCCACCATCTACCCAAATATTAG	2664
DB	341	aAlaGlyAlaThrGlyProArgGlyLeuValGlyGluProGlyProAlaGlySer	359
QY	2663	ACACCAACACAGAAAGCTAGCAATGATTCCCTTCTACTTTGTTAAATAAAGTAA	2604
DB	360	----LysGlyGluThrGlyAsnLysGlyGluProGlySer	371
QY	2603	ATATTTAAATGCCCTGTGTCTCTGTGATGGCAACAGAACCAACAGGCACATCCTGAT	2544
DB	372	-----AlaGlyAlaGlnGlyProProGlyProSerGlyCl	383
QY	2543	AAAAGGTAAGGGGGTGGATCAGCAAAAGACAGTGTCTGT	2500
DB	383	uGluGlyLysArgGlySerProGlyGluProGlySerAlaGlyProAlaGlyProProGl	403
QY	2499	GCTAGGGGA---CCTGGTTCCTGTGTGTGGCCCTCAGGACTCTCCCTACAAATAG	2443
DB	403	yLeuArgGlySerProGlySerArgGlyLeuPro	414
QY	2442	TCATATGTTCAATCCATGGAGGAGTGTTCATCCT	2392



Db	46	-----	:	-----	-ProAlaGly	48
QY	2449	AAATAAGTCATATGTTCAAAATCCCATGGAGGAGTGTTCATCTCTAGAAACTCCCATGCAA	2390			
Db	48	-----	:	-----	-ProAlaGly	48
QY	2389	GAGCTACATTAAACGAAGCTGCAGGTTAAGGGGGTTAGAGATGGGAACACGAGTGACTGA	2330			
Db	49	-----	:	-----	-ProArgGly	57
QY	2329	GTTTATTTCAGCTCCCAAAAACCCCTTCTCTAGGTGTGTCTCAACTAGGAGCTAGCTGTTA	2270			
Db	58	-----	:	-----	-PrometGlyProPro	62
QY	2269	ACCCTGAGCGTGGTAATCCACCTGCGAGAGTCCCGCATTTCCAGTCATCGAGCCCTTCT	2210			
Db	63	-----	:	-----	-GlyProProGlySerProGlyProGlySerPro	77
QY	2209	GGCTCCCTGTATAAGTCCAGACTGAACCCCTTGGAAAGGCTCCAGTCAGC	2156			
Db	78	GlyLeuThrGlyAsnPheAlaAlaGlnTyrSerAspGlyGlyValSerGlyProGly	97			
QY	2155	-----	:	-----	-AGCCCTAGAGACTGGGAGAGAGGAGGAGCGCCAGCGCCACCGCCCA	2111
Db	98	PrometGlyLeuMetGlyProArgGlyProProGlyAlaValGlyAlaProGlyProGln	117			
QY	2110	GCTGTGAGCTACGACCTCAGCAGCACAGGTTGGCAGAGAGCCACATTTACTTTGGC	2051			
Db	118	GlyPheGln	133			
QY	2050	AGCAACAGAACTGGCGGCAGCCGCGCAGCCCATGGGGCTAACAGGAGCGGGAGCTG	1991			
Db	134	ProAlaGlyProArgGlyProAlaGlySerProGlyLysAlaGlyLeuAspGlyHisPro	153			
QY	1990	GGA	1946			
Db	154	GlyLysProGlyArgProGlyGluArgGlyValValGlyProGlnGlyAlaArgGlyPhe	173			
QY	1945	CTACGCTGAGTATTGGCCAACTGCCTTGTCTGTCAATAGTACCTGTGTAGCAAACTAAAT	1886			
Db	174	Pro	186			
QY	1885	GGCGACACACCCAGCCTCGCGCAGACACCATATAGGATGACAGACTG	1835			
Db	187	Gly	197			
QY	1834	-----	:	-----	-GCTGAGCTGGCAATGAGCCCATAAACAGGATGGGCCCACCTG	1790
Db	198	ProGlyAlaGlnGlyValLysGlyGluProGlyAlaProGlyGluAsn	217			
QY	1789	GGACAGCAGGA	1745			
Db	217	LysGlnAlaGlyAlaArgGlyLeuProGlyGluArgGlyArgValGlyAlaProGlyPro	236			
QY	1744	GGCCGGCCCGGAACACCCCTGGCTCGGTGGGCTCACCCACACACCTACGAGAC	1685			
Db	237	AlaGlyAlaArgGlySerAspGlySerValGly	255			
QY	1684	ATCAGGAGGAGGCGCGCAGAGCGCGGTGGAGGTGGGACAGG	1639			
Db	255	eGlySerAlaGlyProProGlyPheProGlyAlaProGlyProLysGlyGluLeuGlyPr	275			
QY	1638	-----	:	-----	-CCACTGCTCCAGCACCCACGTTGCTTACCTAGGGAAG	1592
Db	275	oValGlyAsnProGlyProAlaGlyProAlaGlyProArgGlyGluValGlyLeuProGln	295			
QY	1591	CTTA	1535			
Db	295	LeuSerGlyPro	304			
QY	1534	GTCCCTCGGTATT	1517			

Db 304 yAsnPro-GlyThrAsnGlyLeuThrGlyAlaLysGlyAlaThrGlyLeuProGlyVala 324  
 QY 1516 CAGGAACACCTGCTCTCCCGGTGGTACAGAGGCCA-----GTGCTGA 1472  
 Db 324 laGlyAlaProGlyLeuProGlyProArgGlyIleProGlyProAlaGlyAlaAlaGlyA 344  
 QY 1471 GGGCAGGATCTCAGGCGTGAAGGTCAACCCGGTGAGGCGGTGAAGTGTCAACCAC 1412  
 Db 344 laThrGlyAlaArgGlyLeu-ValGlyGluProGlyProAlaGlySerLysGlyLysSer 363  
 QY 1411 GGGCAC-----ACTGTGGCAGGCGATGTGGCAGCGGACAGCCAGCGGAA 1367  
 Db 364 GlyAsnLysGlyGluProGlySerValGlyAlaGlnGlyProGlyProSerGlyGlu 383  
 QY 1366 AGCTGCCACATGGCCAAATAGACTGCTCGAGTGCCGAATC-----GCTGCCACGACC 1314  
 Db 384 GluGlyLysArgGly-----SerProGlyGluAlaGlySerAlaGlyProAla 399  
 QY 1313 GTTCCATGACAGAGAGAACAGGAGATGGCGCATGCGCCAGGAACAGCCCGGAGTGC 1254  
 Db 400 GlyPro-----ProGlyLeuArgGlySerProGlySerArg---GlyLeu 413  
 QY 1253 CCATCCGAACGCTTCATCATAGTGTCTCCGGGCT-----CGGTGC 1212  
 Db 414 ProGlyAlaAspGlyArgAlaGlyValMetGlyProProGlyAsnArgGlySerThrGly 433  
 QY 1211 CGGGTCACGCTGGGCAGCGCTGTACAGCCCTCGCCAGCAAAATCCGTGTAACA 1152  
 Db 434 ProAlaGlyIleArgGlyProAsnGlyAspAlaGlyArgProGlyGluProGlyLeuMet 453  
 QY 1151 GCGTGAAGTCATAGTGCATCCAGCTGCACAGCTCAGCCAGCAAGAGCGGCGCAGGG 1092  
 Db 454 GlyProArgGlyLeuProGlySerProGlyAsnValGlyProSerGlyLysGluGlyPro 473  
 QY 1091 TGGCGGCATGGCC-----AGCACGCTGGTGCAGCGGG 1056  
 Db 474 ValGlyLeuProGlyIleAspGlyArgProGlyProIleGlyProAlaGlyProArgGly 493  
 QY 1055 GAAGCAGGCGCCAGGTTCCGGAAGCAAGCGGCGCGCATGACACAGCAGTGGGGCG 996  
 Db 494 GluAlaGlyAsnIleGlyPheProGlyProLysGlyProSerGlyAspProGlyLysPro 513  
 QY 995 ACAAGGAGCGCGGACAGCCCTTCGTGCTGGTGGTGGGCGCCAGCGCTG-----945  
 Db 514 GlyGluArgGlyHisProGlyLeuAlaGlyAlaArgGlyAlaProGlyProAspGlyAsn 533  
 QY 944 -----CCTCTCAGCCAGCAGCAGTGTGCTGTCTAGCAGGTGAGGAAGA 900  
 Db 534 AsnGlyAlaGlnGlyProProGlyProGlnGlyValGln-----546  
 QY 899 TGAGGCTGACAGCGCCAAAGAGGCACT-----CCTCTGGGTGGCCAGGTAGGGGGCCA 846  
 Db 547 ---GlyGlyLysGlyGluGlnGlyProAlaGlyProProGlyPheGlnGly-----562  
 QY 845 GGGCAGTGGTGTCCAGCTCAATGGCAGGAGGAGGTAGCCAGGAGCCCGCCAGACTGA 786  
 Db 563 -----LeuProGlyProSerGlyThrGlyGluValGlyLysProGlyGlu-----578  
 QY 785 TCATGAAGCATAGACAGATAGCTTGGGACAGAGTGGT---CCGGGTCCCGGAAGAGT 729  
 Db 579 -----ArgGlyLeuProGlyGluPheGlyLeuProGlyPro-----590  
 QY 728 CAGAGACAGGCGCTCCAGTGGAGTGAAGCACACCTGGCCACAGAGTCCAGAGCCCA 669  
 Db 591 -----AlaGlyProArgGlyGluArgGlyThrProGlyLysGlyAlaAlaGlyPro 608  
 QY 668 -----CGCCAGGATGACAGTGCAGCTCA-----642  
 Db 609 SerGlyProIleGlySerArgGlyProSerGlyAlaProGlyProAspGlyAsnLysGly 628  
 QY 641 -----GGGGCTGGGATCCGGCAGCAGCAGCCCTGTCTAGCCAGCCGCGCC 597  
 Db 629 GluAlaGlyAlaValGlyAla-----ProGlySerAlaGlyAlaSerGlyProGlyGly 646

QY 596 TTGGGATGAGAAAGAGCTCAGCAGGATGCCCA-----564  
 Db 647 LeuProGlyGluArgGlyAlaAlaGlyIleProGlyGlyLysGlyGluLysGlyGluThr 666  
 QY 563 -----AGGACAGTCCAGCATCA 546  
 Db 667 GlyLeuArgGlyAspThrGlyAsnThrGlyArgAspGlyAlaArgGlyIleProGly-----685  
 QY 545 AGGGCGGCGGCGCATAGCTCCAGCGCAGTGGTCACTGGCTGAGCCTAGGAGCGGGA 486  
 Db 686 AlaValGlyAlaProGlyProAlaGlyAlaSerGlyAspArgGlyGluAlaGlyAlaAla 705  
 QY 485 CACAGACAGGCGGCGCAGTGGACCAATGCCAGCAGCAGTGTCTCTCTCTA 426  
 Db 706 GlyProSerGlyProAla-----711  
 QY 425 CCCCCTCCAGCAGCAGGCGGCACATAGGTGCTGCGCGCAACACACACCTCCA 366  
 Db 712 -----GlyProArgGlySerPro 717  
 QY 365 GGCCAAAG-----TTAGCAGGTTCAGCAGCAAGA 336  
 Db 718 GlyGluArgGlyGluValGlyProAlaGlyProAsnGlyPheAlaGlyProAlaGlyAla 737  
 QY 335 GCTGGGCTTCCGGTGC-----318  
 Db 738 AlaGlyGlnProGlyAlaLysGlyGluLysGlyThrLysGlyProLysGlyGluAsnGly 757  
 QY 317 -----CGACAGCGGCTCACCACAGCTCTGGACCATAGTGGGCGCAGCGCG-----269  
 Db 758 IleValGlyProThrGlySerValGlyAlaAlaGlyPro-SerGlyProAsnGlyProPr 777  
 QY 268 -----TAGGCTC 261  
 Db 777 OglyProValGlySerArgGlyAspGlyGlyProProGlyMetThrGlyPheProGlyAl 797  
 QY 260 AGGGGCGGTTCAAGCACTCCAGAACTGCTGCTCGGCTCTGCTCCAGAACTGGCGC 201  
 Db 797 aAlaGlyArgThrGlyProPro-----GlyProSerGlyIleAl 810  
 QY 200 CTCCTCTCTGCTGCGCCCAACTGCTAGGAATCAGCCAGGCGCCCATTTTCTCCAGCC 141  
 Db 810 aGlyProPro-----GlyProProGlyAla-----818  
 QY 140 CTTTGGTCCCGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGGGCGACCTCA 81  
 Db 819 -AlaGly-LysGluGlyIleArgGlyProArgGlyAspGlnGlyProValGlyArg---t 837  
 QY 80 GTGGGACAGCTCTCATCTCAGATCCTGCG-----CGAGGCGCGGCTGT 33  
 Db 837 hrGlyGluThrGlyAlaSerGlyProGlyPheValGlyGlyLysGlyProSerGlyG 857  
 QY 32 CACCGGAGCGCAGC 19  
 Db 857 LuProGlyThrAla 861

## RESULT 41

## CA2B\_HUMAN

## ID

## CA2B\_HUMAN

## STANDARD;

## PRT; 1736 AA.

## AC

## P13942; Q13271; Q13272; Q07751; Q99866; Q9UIP9;

## DT

## 01-JAN-1990 (Rel. 13, Created)

## DT

## 16-OCT-2001 (Rel. 40, Last sequence update)

## DT

## 15-JUN-2002 (Rel. 41, Last annotation update)

## DE

## Collagen alpha 2(XI) chain precursor.

## GN

## COL11A2.

## OS

## Homo sapiens (Human).

## OC

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## OC

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homq.

## OX

## NCBI\_TaxID=9606;

## RN

## SEQUENCE FROM N.A.

## RP

## MEDLINE=96032717; PubMed=7559422;

## RX

RA Vuoristo M.M., Pihlajamaa T., Vandenbergh P., Prockop D.J.,  
 RA Ala-Kokko L.;  
 RT "The human COL1A2 gene structure indicates that the gene has not  
 RT evolved with the genes for the major fibrillar collagens.";  
 RL J. Biol. Chem. 270:22873-22881(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tubby B.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 59-807 FROM N.A.  
 RA TISSUE-Cartilage;  
 RC MEDLINE=93314796; PubMed=8325374;  
 RX Zhidkova N.I., Brewton R.G., Wayne R.;  
 RA "Molecular cloning of PARP (proline/arginine-rich protein) from human  
 RT cartilage and subsequent demonstration that PARP is a fragment of the  
 RT NH2-terminal domain of the collagen alpha 2(XI) chain.";  
 RL FEBS Lett. 326:25-28(1993).  
 RN [4]  
 RP SEQUENCE OF 730-1690 FROM N.A.  
 RX MEDLINE=89340485; PubMed=2760050;  
 RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,  
 RA van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;  
 RT "The human alpha 2(XI) collagen (COL1A2) chain. Molecular cloning of  
 RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen  
 RT with differences in genomic organization.";  
 RL J. Biol. Chem. 264:13910-13916(1989).  
 RN [5]  
 RP SEQUENCE OF 1-537 FROM N.A.  
 RX MEDLINE=96435918; PubMed=8838804;  
 RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;  
 RT "The human alpha 2(XI) collagen gene (COL1A2): completion of coding  
 RT information, identification of the promoter sequence, and precise  
 RT localization within the major histocompatibility complex reveal  
 RT overlap with the KES gene.";  
 RL Genomics 32:401-412(1996).  
 RN [6]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=95238468; PubMed=7721876;  
 RA Zhidkova N.I., Justice S.K., Wayne R.;  
 RT "Alternative mRNA processing occurs in the variable region of the  
 RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";  
 RL J. Biol. Chem. 270:9486-9493(1995).  
 RN [7]  
 RP DISEASE.  
 RX PubMed=10677296;  
 RA Melkonian M., Brunner H.G., Manouvier S., Hennekam R.,  
 RA Superti-Furga A., Kaeerlaeinen H., Pauli R.M., van Essen T.,  
 RA Warman M.L., Bonaventure J., Miny P., Ala-Kokko L.;  
 RT "Autosomal recessive disorder otospondyloymeagapiphyseal dysplasia is  
 RT associated with loss-of-function mutations in the COL1A2 gene.";  
 RL Am. J. Hum. Genet. 66:368-377(2000).  
 RN [8]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97255959; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [9]  
 RP VARIANT OSMED ARG-661.  
 RX MEDLINE=95163096; PubMed=7859284;  
 RA Vikkula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,  
 RA Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,  
 RA van den Hoogen F.H.J., Ropers H.-H., Wayne R., Cheah K.S.E.,  
 RA Olsen B.R., Warman M.L., Brunner H.G.;  
 RT "Autosomal dominant and recessive osteochondrodysplasias associated  
 RT with the COL1A2 locus.";  
 RL Cell 80:431-437(1995).  
 RN [10]  
 RP VARIANTS GLY-593; LYS-824; LEU-879; THR-1316 AND GLN-1600.  
 RX PubMed=9585596;

RA Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,  
 RA Yone K., Matsunaga S., Otterud B., Inoue I., Leppert M.;  
 RT "Genetic mapping of ossification of the posterior longitudinal  
 RT ligament of the spine.";  
 RL Am. J. Hum. Genet. 62:1460-1467(1998).  
 RN [11]  
 RP VARIANT WZS GLU-955.  
 RX PubMed=9805126;  
 RA Pihlajamaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,  
 RA Giedion A., Wiesbauer P., Spranger J., Ala-Kokko L.;  
 RT "Heterozygous glycine substitution in the COL1A2 gene in the original  
 RT patient with the Weissenbacher-Zweymueller syndrome demonstrates its  
 RT identity with heterozygous OSMED (nonocular Stickler syndrome).";  
 RL Am. J. Med. Genet. 80:115-120(1998).  
 RN [12]  
 RP VARIANT STL3 940-GLY--PRO-948 DEL.  
 RX PubMed=9506662;  
 RA Sirko-Osadsa D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,  
 RA Robin N.H.;  
 RT "Stickler syndrome without eye involvement is caused by mutations in  
 RT COL1A2, the gene encoding the alpha-2(XI) chain of type XI  
 RT collagen.";  
 RL J. Pediatr. 132:368-371(1998).  
 RN [13]  
 RP VARIANTS DENA13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.  
 RX MEDLINE=20047768; PubMed=10581026;  
 RA McGurt W.T., Prasad S.D., Griffiths A.J., Kunst H.P.M., Green G.E.,  
 RA Spargel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E.,  
 RA King M.-G., Brunner H.G., Cremers C.W.R.J., Takanosu M., Li S.-W.,  
 RA Arita M., Wayne R., Prockop D.J., Van Camp G., Smith R.J.H.;  
 RT "Mutations in COL1A2 cause non-syndromic hearing loss (DFNA13).";  
 RL Nat. Genet. 23:413-419(1999).  
 CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD  
 CC OF ALPHA 3(XI)-1(II).  
 CC -!- ALTERNATIVE PRODUCTS: 8 isoforms; 1 (shown here), 2, 3, 4, 5, 6, 7  
 CC and 8; may be produced by alternative splicing. They lack exons 6,  
 CC 7 or 8 or a combination of these exons.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- PTM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH  
 CC PROTEIN OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING  
 CC EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE  
 CC CARTILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT  
 CC AMOUNTS.  
 CC -!- DISEASE: Defects in COL1A2 are the cause of Stickler syndrome  
 CC type 3 (STL3). It is an autosomal dominant disorder characterized  
 CC by oro-facial, auditory and skeletal manifestations, such as  
 CC midfacial hypoplasia, cleft palate, osteoarthritis, and  
 CC sensorineural hearing loss. Differently from Stickler syndrome  
 CC type 1 and 2, no ocular involvement is observed. This disorder is  
 CC also referred to as Stickler-like syndrome or non-ocular Stickler  
 CC syndrome.  
 CC -!- DISEASE: Defects in COL1A2 are the cause of autosomal recessive  
 CC otospondyloymeagapiphyseal dysplasia (OSMED), a skeletal dysplasia  
 CC accompanied by severe hearing loss. The phenotype overlaps that of  
 CC autosomal dominant skeletal disorders (Stickler and Marshall  
 CC syndromes) but can be distinguished by disproportionately short  
 CC limbs and lack of ocular involvement.  
 CC -!- DISEASE: Defects in COL1A2 are the cause of Weissenbacher-  
 CC zweymueller syndrome (WZS), an autosomal dominant disorder allelic  
 CC with STL3 and OSMED. It is also referred to as heterozygous OSMED.  
 CC -!- DISEASE: Defects in COL1A2 are the cause of autosomal dominant  
 CC nonsyndromic sensorineural deafness type 13 (DFNA13). Affected  
 CC individuals experience progressive hearing loss beginning in the  
 CC second to fourth decades, eventually making use of amplification  
 CC mandatory.  
 CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -!- DATABASE: NAME-Hereditary hearing loss homepage;  
 CC NOTE=Gene page;

647	Db	----	ProProGlyGlnGlnGlyThrProGlyThr-----	GlnG 658	
2690	QY	GGATCCCCCACCCTACCCCAATATTAGACACACACACAAAGCT-----	2645		
658	Db	lyLeuProGlyProGlnGlyAlaIleGlyProHisGlyGluLysGlyProGlnGlyLysP 678			
2644	QY	-----	AGCAATGGAT 2635		
678	Db	roGlyLeuProGlyMetProGlySerAspGlyProProGlyHisProGlyLysGluGlyP 698			
2634	QY	TCCCTTCTACTTTGTTAAATAAATAAGTTAAATATTAAATGCTGTGCTCTGTGATGG 2575			
698	Db	roProGlyThrLysGlyAsnGln-----	GlyProSerG 709		
2574	QY	CAACAGAGGACCAACAGGCGACATCCTGATAAAGGTAAAGAGGGGGTGGATCAGCAA 2515			
709	Db	lyProGlnGlyProLeuGlyThrProGlyValLysGlyValAspGlyLeuA 729			
2514	QY	AAGACAGTGTGTGGCTGAGGGGACCTGGTCTTGTGTGTGGCTCAGGACTCTTCC 2455			
729	Db	rg-----	GlyLeuLysG 733		
2454	QY	CCTACAATAAATCATATGTTCAATCCCATGGAGGAGTCTTTCATCTCTAGAAAACCTCCA 2395			
733	Db	lyHisLys-----	GlyGluLysGlyGluAspGlyPhe-----	ProG 745	
2394	QY	TGCAAGAGGACTACATTAAACGAAGCTCAGAGTAAAGGGGCTTAGAGAT-----	GGGAAC 2341		
745	Db	lyPheLys-----	GlyAspIleGlyValLysGlyAspArgGlyGluValGlyValP 762		
2340	QY	CAGGTAGTCAAGTTATTTCAGTCCCAAAACCT-----	2306		
762	Db	roGlySerArgGlyGluAspGlyProGluGlyProLysGlyArgThrGlyProThrGlyA 782			
2305	QY	-----	TCTCTAGGTGTGTCTCAACTAGGAG 2281		
782	Db	spProGlyProProGlyLeuMetGlyGluLysGlyLysLeuGlyValProGlyLeuPro- 801			
2280	QY	GCTAGCTGTTAACCCCTGAGCCTGAGCTGATCCACCTGCAGAGTCCCGCATCCAGTGCAT 2221			
802	Db	-----	GlyTyrProGlyArgGlnGlyProLysGlySerLeuG 814		
2220	QY	GGAGCCCTCTGGCCTCCCTGTATAGTCCAGAGCTGAAACCCCTTGGAGGCCCTCCAGT 2161			
814	Db	lyPhePro-----	AlaSerGlyGluLysGlyValAlaArgGlyLeu---S 831		
2160	QY	CAGCAGCCCTAGAGACTGGGGAGAGAGGAGGAGCCGCCGCCGCCAGCTGTGCAGC 2101			
831	Db	erGlyLysSerGlyProArgGlyGluArg-----	GlyProThrGlyProArgGlyGlnA 849		
2100	QY	TACGCACCTCAGCAGCAGCAGGGTGGCAGCAGAGCCACATTACTTTGGCAGCAACAGAA 2041			
849	Db	rgGlyProArgGlyAlaThrGlyLysSerGlyAlaLysGlyThrSerGlyGlyAsp-----	867		
2040	QY	ACTGGCGCCAGCCGCGAGCCCATGGGCTTAACAGGAGCGGGAGCTGGGA---CCCA 1984			
868	Db	-----	GlyProHisGlyPro-----	ProGlyGluArgGlyLeuProG 880	
1983	QY	GTGAGGCGAGCCCTCCACCCCAATGTGCTGGAGTTTCTACGCTGAGTATTGTGCCCAAG 1924			
880	Db	lyProGlnGlyPro-----	884		
1923	QY	TGCTCTTGTCAATATACTACCTGTGTAGCAAGTAAATGGCA-----	188		
885	Db	-----	AsnGlyPheProGlyProLysGlyP 893		
1880	QY	---CCAGACCCA-----	GGCCTCGCGCAGACACCAATATAGGCAGTGACAGACTG 1833		
893	Db	roProGlyProProGlyLysAspGlyLeuProGlyHisProGlyGlnArgGlyGluValG 913			
1834	QY	GCTGAGCTGGACAATGGAGCCCATAAACAGG-----	ATGGGCCACCTGGGA 178		
913	Db	lyPheGlnGlyThrGlyPro-----	ProGlyValValGlyProGlnGlyA 932		

Alignment Scores:	6.6e-07	Length:	1736
Pred. No.:	301.00	Matches:	317
Score:	30.91%	Conservative:	61
Percent Similarity:	30.91%	Mismatches:	385
Best Local Similarity:	25.92%	Indels:	462
Query Match:	4.84%	Gaps:	76
DB:	1		
US-09-759-143-110 (1-3410) x CA2B_HUMAN (1-1736)			
QY	3035	CTGGTGGGGAAGTTGGGGTAGGGGAAGTTGGGGTAGGGGAATTTTGGCAGTGCC	2976
		:               :	
Db	566	LeuProGlyGluGlySclHisArgGlyAspThrGlyAlaGlnGlyLeuProGly-----	583
QY	2975	TTCATCAGCCAGTCCTCAGACAGAGTAGAGGGGAGTGGGAAGTGGGGGAGCAACGAGCTGGG	2916
Db	584	-----ProProGlyGlu-----AspGlyGluArgGlyAspAspGlyGluIleGly	598
QY	2915	CCAAGA-----GAAGAGGGTGGTTAGGGAAGCCGTTGACACTGAGCCCA	2868
Db	599	ProArgGlyLeuProGlyGluSerGlyProArgGlyLeu---LeuGlyProGlyGlyPro	617
QY	2867	CCCTCTACCTTCCTTCAACACCCCTAACCTTTGGGTAACAGCATTTGGAATATCATTTGGG	2808
Db	618	ProGly--lleProGlyProPro-----GlyValArgGlyMeta	630
QY	2807	ATGAGTAGAATTTCCAAGGTCCTGGGTTAGGCATTTTGGGGGCCAG---ACCCAGGAG	2751
Db	630	sp-----GlyProGlnGlyPro---LysGlySerLeuGlyProGlnGlyGluProGly-	646
QY	2750	AAGAGATTCTGGCAATGATCAGCCCAATGACCAAGCTATCTCAGGGGAGACTGATTGTG	2691



QY	1786	CAGCAGGAGGCACTATTCCA	---GGATGCCGAGGTCTCCAGGCAGATGCCCGCCGCGAA	1731
DB	932	laAlaGlycylThrGlyProMetGlyGluArgGlyHis	-----ProGlyProProG	949
QY	1730	--CCACCCTGGCCTCGGTGGCTCACCACCCACCACACGTCACGAGACATCACAGGCAG	1673	
DB	949	lyPro-ProGlyGluGlnGlyLeu-ProGlyThrAlaGlyLysGluGlyThrLysGlyAs	968	
QY	1672	G-----GCCCGCCAGACGCGGTGGAGTGGGAGCAGGCCACTGCCTCC	1628	
DB	968	pProGlyProProGlyAlaProGlyLysAspGlyProAlaGlyLeuArgGlyPheProGly	988	
QY	1627	AGCACCCAGTGTCCATTAGGAAGGAGTCCAGGCTTA	1588	
DB	988	yGluArgGlyLeuProGlyThrAlaGlyGlyProGlyLeuLysGlyAsnGluGlyProse	1008	
QY	1587	-----GGGCTGGCAGGAAC	1556	
DB	1008	rglyProProGlyPro-AlaGlySerProGlyGluArgGlyAlaAlaGlySerGlyGlyp	1028	
QY	1555	CTCACTGCTAGCACTCCAGTGTCCCTCGGTATTGGGCAGGAACACCTGCTTCCTCCCG	1496	
DB	1028	ro-----	1036	
QY	1495	GTGTTAGAGGAGGCCAGTGTGTAGGCCAGGATCTGCAGGCTGAGAAGGTGAACCCGT	1436	
DB	1036	lyProGlnGlyProPro-----	1051	
QY	1435	GAGGGCGGTGAAGTGTCTACCACGCCACACTGTGGCAGCAGCATGTGCACCGGCAGC	1376	
DB	1052	GluLysGly-ProIleGlyProThrGlyArgAspGlyVal	1067	
QY	1375	CACAGGAAAGCTGCCACACTGCCCAATAGACTGCTCGAGTCCCAATCGTGCACCAG	1316	
DB	1067	roValGly-----	1074	
QY	1315	CCGGTCCATGACCAGAGAGAAGACCAGGAGATGGCGCACTGCAGGAACA	1266	
DB	1074	laglyPro-----	1086	
QY	1265	-----GCCCAGGCTGCCCATCCGACGCCTTCATCATAGTGTCTCCGGGC	1220	
DB	1086	ysGlyGluValGlyAspProGlyGlnLysGlyThrLysGlyAsnLysGlyGluHisGlyp	1106	
QY	1219	CTCGTGGCGGCTCAGCTCTGGGCACGCCCTGTGTACGCCCTCGCCACCAATCCGT	1160	
DB	1106	ro-----	1115	
QY	1159	GTAAACAGCGTGAAGGTCATGAGTCCATCCAGCTGCACACTGCACCAAGAGCGG	1100	
DB	1116	-----	1123	
QY	1099	GCACAGG-----TCCGGGGCATGCGGCAGCACAGCT-----	1052	
DB	1123	lyAlaaspGlyGluProGlyAlaArgGlyProGlnGlyHisPheGlyAlaLysGlyAspG	1143	
QY	1051	CAGGGCGCCAGGTTCCGGAAGCCAAGCGGGCCC	1005	
DB	1143	luGlyThrArgGlyPheAsnGlyProProGlyProIleGlyLeuGlnGlyLeuProGlyp	1163	
QY	1004	--AGTGGGCGCAAGAGGGGGCCAGACGCCCTTCTGCTGCTCGGTGGGGCCACGGC	947	
DB	1163	roserGly-----	1177	
QY	946	TGCTTCCTCAGCCACGAGTGTGGCTGCTACGCAAGTGAAGAGATGAGGTGAGCAG	887	
DB	1177	lyProProGlyProPro-----	1189	
QY	886	GCCAAAGAGGCACT-----	845	
DB	1189	lyProAsnGlyAlaaspGlyProGlnGlyProProGlyGlyValGlyAsnLeuGlyPro	1208	

QY	844	GGCACTGGTGTCCCATGTCATAGCGCAGGCAGGAGGTAGCCACGACGAGGCCCCCACAAAGACTGAT	705
DB	1209	-----ProGlyGluLysGlyGluProGlyGlySerPro	1221
QY	784	CATGAAGGCATAGACAGACTAGCCCTGGCGGACAGTGCTCCGGGTCCCGGAAGAGGTCAGA	725
DB	1222	-----GlyIleGlnGlyProGlyValLysGlyProArgGlyGlyLeuArgGlyGluL	1239
QY	724	GA-----GCAGGGCTCCAGTGAGTGAAGCACACCTTGGCCACAGAAAGTCCACGACGCC	671
DB	1239	yseLyGluSerGlyGlnProGlyGluProGlyProGlyAlaLysGlyProGlnGlyA	1259
QY	670	CA-----CGCCAGGATGACGAGTGCCAGCTCCAGGGCCCTGGGATCCGG	636
DB	1259	sPAspGlyProLysGlyAsnProGly-----ProValGlyPheProGly-----AspProg	1276
QY	625	GCACAGCAGCCCTGCTACCAGCCGG-----	600
DB	1276	lyProProGlyGluGlyGlyProArgGlyGlnAspGlyAlaLysGlyAspArgGlyGluA	1296
QY	599	-----C	599
DB	1296	spGlyGluProGlyGlnProGlySerProGlyProThrGlyGluAsnGlyProProGlyP	1316
QY	598	CCTTTGGATGAAAAGGCTCACGAGATGCCAACGACAGTAGTCCCCAGATGAAGGCCG	539
DB	1316	roLeuGly--LysArgGlyProAlaGlySerProGlySerGlyGlyArg-----GlnGlyG	1334
QY	538	GCGGGCGGCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGGGGACACAGAC	479
DB	1334	lyLysGly-----AlaLysGlyAspProGlyAlaIleGlyAlaProGlyLysT	1350
QY	478	CAGGCCACGACACTGGACAATGGCCACGACCATGGTCATCACTTCTCTTACCCCCAC	419
DB	1350	hrGlyProValGlyProAlaGlyProAlaGlyLysProGlyProAspGlyLeuArgGlyL	1370
QY	418	TTCACGACGACAGCGCGCACATAGGTGATGCTCGTGGCGGCCAAACACA-----	372
DB	1370	eUProGlySerValGlyGlnGly-----ArgProGlyAlaThrGlyGlnAlaG	1387
QY	371	--CCTCCACGAGCAAAGTTAGCAGGTGACACGACAGACTGGCTTCCGTTGCCGCGAG	314
DB	1387	lyProProGlyPro-----valGlyProProGlyLeuProGlyLeuArgGlyAspAlaG	1405
QY	313	CAGGC-----	309
DB	1405	lyAlaLysGlyGluLysGlyHisProGlyLeuIleGlyLeuIleGlyProProGlyGluG	1425
QY	309	-----	309
DB	1425	lNGlyGluLysGlyAspArgGlyLeuProGlyProGlnGlySerProGlyGlnLysGlyG	1445
QY	308	-----GGCTCACCCACAGCCTCTGGACCATAGTGGGCCACGCGGG-----	269
DB	1445	luMetGlyIleProGlyAlaserGlyPro-IleGlyProGlyGlyProProGlyLeuPro	1464
QY	268	-----TAGGGCTCAGGGGCGGCTTCAGGCAC	242
DB	1465	GlyProAlaGlyProLysGlyAlaLysGlyAlaThrGlyProGlyGlyProLysGlyGlu	1484
QY	241	CCAGAACTGCTGCTCGGCTCTGCTCCAGAAAGTGGCGGCTCTCTCTCTCT-----TGC	188
DB	1485	LysGlyVal-----GlnGlyProProGlyHisProGlyProProGlyGluVal	1500
QY	187	TGGCGCCAACCTGCTAGGAATCAGCCA-----GGGCGCCCAT	152
DB	1501	IleGlnProLeuProIleGlnMetProLysLysThrArgSerValaspeLySerArg	1520
QY	151	TTCTGCCAG-----CCCTTTGGTCCCGGTCACGCTTCTCACCCATGCTC	107
DB	1521	LeuMeGlnGluAspGluAlaIleProThrGlyGlyAlaProGlySerProGlyGlyLeu	1540
QY	106	AACACCTGCTGCTGTGGGACACCTCAGTGGGACACGCTCTCATCACTCAGAG56	

Db 1541 GluGlutle-----PheGlySerLeuAspSerLeuArg 1551  
 RESULT 42  
 CA334\_HUMAN  
 ID CA334\_HUMAN STANDARD; PRT; 1670 AA.  
 AC Q01955; Q9BQ02;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).  
 GN COL4A3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94364994; PubMed=8083201;  
 RA Maruyama M., Leinonen A., Mochizuki T., Tryggvason K., Readers S.T.;  
 RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
 RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
 RT human tissues.";  
 RL J. Biol. Chem. 269:23013-23017(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Leinonen A.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., VARIANTS AS E-297; R-407; R-640; R-1167; E-1207;  
 RP Q-1215; S-1277; T-1330; E-1334; E-1347 AND C-1661, AND VARIANTS R-43;  
 RP E-162; Y-326; H-408; R-451; L-574; E-1269 AND P-1474.  
 RX MEDLINE=21064696; PubMed=11134255;  
 RA Heidt L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,  
 RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
 RT autosomal Alport syndrome.";  
 RL J. Am. Soc. Nephrol. 12:97-106(2001).  
 RN [4]  
 RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=93015826; PubMed=1400291;  
 RA Quinones S., Bernal D., Garcia-Soto M., Elena S.F., Saus J.;  
 RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
 RT Goodpasture antigen (alpha 3(IV)NCI). Identification of a potentially  
 RT antigenic region at the triple helix/NCI domain junction.";  
 RL J. Biol. Chem. 267:19780-19784(1992).  
 RN [5]  
 RP SEQUENCE OF 1453-1670 FROM N.A.  
 RX MEDLINE=91353570; PubMed=1882840;  
 RA Morrison K.E., Maruyama M., Yang-Feng T.L., Readers S.T.;  
 RT "Sequence and localization of a partial cDNA encoding the human alpha  
 RT 3 chain of type IV collagen.";  
 RL Am. J. Hum. Genet. 49:545-554(1991).  
 RN [6]  
 RP SEQUENCE OF 1331-1670 FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=92147878; PubMed=1737849;  
 RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
 RA Pusey C.D.;  
 RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
 RT to be the alpha 3 chain of type IV collagen.";  
 RL J. Clin. Invest. 89:592-601(1992).  
 RN [7]  
 RP SEQUENCE OF 1644-1670 FROM N.A.  
 RC TISSUE=Kidney;  
 RA Ding J.;  
 RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94124597; PubMed=8294492;  
 RA Feng L., Xia Y., Wilson C.B.;  
 RT "Alternative splicing of the NCI domain of the human alpha 3(IV)  
 RT collagen gene. Differential expression of mRNA transcripts that  
 RT predict three protein variants with distinct carboxyl regions.";  
 RL J. Biol. Chem. 269:2342-2348(1994).  
 RN [9]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=98196854; PubMed=9537506;  
 RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshiooka H.,  
 RA Ninomiya Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
 RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
 RT 2q36.";  
 RL FEBS Lett. 424:11-16(1998).  
 RN [10]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=93280184; PubMed=8505332;  
 RA Bernal D., Quinones S., Saus J.;  
 RT "The human mRNA encoding the Goodpasture antigen is alternatively  
 RT spliced.";  
 RL J. Biol. Chem. 268:12090-12094(1993).  
 RN [11]  
 RP VARIANT PRO-1474.  
 RX MEDLINE=95078827; PubMed=7987301;  
 RA Lemink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
 RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,  
 RA Readers S.T., Smeets H.J.M.;  
 RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
 RT recessive Alport syndrome.";  
 RL Hum. Mol. Genet. 3:1269-1273(1994).  
 CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/  
 CC NIDOGEN.  
 CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/V AND  
 CC 3/L5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR  
 CC C-TERMINAL NCI DOMAINS.  
 CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,  
 CC COCHLEA, LUNG AND BRAIN.  
 CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NCI) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG-CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL  
 CC N-LINKED GLYCOSYLATION SITE.  
 CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.  
 CC -!- DISEASE: ANTIBODIES AGAINST THE NCI DOMAIN OF ALPHA3(IV) MEDIATE  
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS  
 CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.  
 CC -!- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I  
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY  
 CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,  
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN  
 CC MALES AND FEMALES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC  EMBL; X80031; CAA56335.1;
DR  EMBL; AJ288487; CAC36101.1; JOINED.
DR  EMBL; AJ288488; CAC36101.1; JOINED.
DR  EMBL; AJ288489; CAC36101.1; JOINED.
DR  EMBL; AJ288490; CAC36101.1; JOINED.
DR  EMBL; AJ288491; CAC36101.1; JOINED.
DR  EMBL; AJ288492; CAC36101.1; JOINED.
DR  EMBL; AJ288493; CAC36101.1; JOINED.
DR  EMBL; AJ288494; CAC36101.1; JOINED.
DR  EMBL; AJ288495; CAC36101.1; JOINED.
DR  EMBL; AJ288496; CAC36101.1; JOINED.
DR  EMBL; AJ288497; CAC36101.1; JOINED.
DR  EMBL; AJ288498; CAC36101.1; JOINED.
DR  EMBL; AJ288499; CAC36101.1; JOINED.
DR  EMBL; AJ288500; CAC36101.1; JOINED.
DR  EMBL; AJ288501; CAC36101.1; JOINED.
DR  EMBL; AJ288502; CAC36101.1; JOINED.
DR  EMBL; AJ288503; CAC36101.1; JOINED.
DR  EMBL; AJ288504; CAC36101.1; JOINED.
DR  EMBL; AJ288505; CAC36101.1; JOINED.
DR  EMBL; AJ288506; CAC36101.1; JOINED.
DR  EMBL; AJ288507; CAC36101.1; JOINED.
DR  EMBL; AJ288508; CAC36101.1; JOINED.
DR  EMBL; AJ288509; CAC36101.1; JOINED.
DR  EMBL; AJ288510; CAC36101.1; JOINED.
DR  EMBL; AJ288511; CAC36101.1; JOINED.
DR  EMBL; AJ288512; CAC36101.1; JOINED.
DR  EMBL; AJ288513; CAC36101.1; JOINED.
DR  EMBL; AJ288514; CAC36101.1; JOINED.
DR  EMBL; AJ288515; CAC36101.1; JOINED.
DR  EMBL; AJ288516; CAC36101.1; JOINED.
DR  EMBL; AJ288517; CAC36101.1; JOINED.
DR  EMBL; AJ288518; CAC36101.1; JOINED.
DR  EMBL; AJ288519; CAC36101.1; JOINED.
DR  EMBL; AJ288520; CAC36101.1; JOINED.
DR  EMBL; AJ288521; CAC36101.1; JOINED.
DR  EMBL; AJ288522; CAC36101.1; JOINED.
DR  EMBL; AJ288523; CAC36101.1; JOINED.
DR  EMBL; AJ288524; CAC36101.1; JOINED.
DR  EMBL; AJ288525; CAC36101.1; JOINED.
DR  EMBL; AJ288526; CAC36101.1; JOINED.
DR  EMBL; AJ288527; CAC36101.1; JOINED.
DR  EMBL; AJ288528; CAC36101.1; JOINED.
DR  EMBL; AJ288529; CAC36101.1; JOINED.
DR  EMBL; AJ288530; CAC36101.1; JOINED.
DR  EMBL; AJ288531; CAC36101.1; JOINED.
DR  EMBL; AJ288532; CAC36101.1; JOINED.
DR  EMBL; AJ288533; CAC36101.1; JOINED.
DR  EMBL; AJ288534; CAC36101.1; JOINED.
DR  EMBL; AJ288535; CAC36101.1; JOINED.
DR  EMBL; AJ288536; CAC36101.1; JOINED.
DR  EMBL; AJ288537; CAC36101.1; JOINED.
DR  EMBL; AJ288538; CAC36101.1; JOINED.
DR  EMBL; M92993; AAA21610.1; -
DR  EMBL; M55790; AAB19637.1; -
DR  EMBL; M81379; AAA51556.1; -
DR  EMBL; L08650; AAA52044.1; -
DR  EMBL; U02519; AAA18942.1; -

Alignment Scores:
Pred. No.: 6,97e-07
Score: 300.50
Percent Similarity: 28.31%
Best Local Similarity: 23.90%
Query Match: 4.83%
DB: 1

US-09-759-143-110 (1-3410) x CA34_HUMAN (1-1670)
Oy 3068 GGTCTCCAGTAGCTCAACAGGGTGTGGAGAGTGTGGGAAAGTGTGGGGTAGGGGA 3009
Db 76 GlyProLysGlyPheProGly-----LeuProGlyLeuThrGlySerLysGly 91

```

```

Oy 3008 AAGTTGGGGTAGGGGAAATTTGGCAGTCGCTTCATCAGCCAGTCCTAGAGAGTA 2949
Db 92 ValArgGlyIleSerGlyLeuProGlyPheSerGly---SerProGlyLeuProGlyThr 110
Oy 2948 GAGGGG-----AGTGGAAAGTGGGGG 2928
Db 111 ProGlyAsnThrGlyProTyrGlyLeuValGlyValProGlyCysSerGlySerLysGly 130
Oy 2927 AACCAAGGCTGGGCCAAGAGAAGAGGGGTGGTGGGAAGCCGTTGAGACCTGAAGCCCA 2868
Db 131 GluGlnGlyPhePro-----GlyLeuProGlyThrProGlyTyProGlyIlePro 147
Oy 2867 CCCTCTACCTTCCTTCAA-----CACCTAACCTTGGGTAAACACATTTGGAA 2820
Db 148 GlyAlaAlaGlyLeuLysGlyGlnLysGlyAlaPro-AlaLysGlyGluAspIleGluLe 167
Oy 2819 TTATCATTTGGGATGAGTAGAATTTCAAGGTCCTGGGTAGGCATTTTGGGGGCCAGA 2760
Db 167 uAspAlaLysGlyAspProGlyLeuPro-----GlyAlaProGly 180
Oy 2759 CCCAGGAGAGAAGATTCTGGCAATGATCAGCCCAATGACCACTATCTCAGGGGACCT 2700
Db 180 yPro-----GlnGlyLeuProGlyProGlyPheProGlyPro-- 193
Oy 2699 GATTGTTGGGATCCCCACCCCTACCCAAATATTTAGACACCAACACAGAAAGCTACAA 2640
Db 194 ----ValGlyProProGlyPro-----ProGlyPhePh 203
Oy 2639 TGGATTCCTCTACTTTGTTAAATAATAAGTTAAATATTAAATGCCGTGCTCTGT 2580
Db 203 eGlyPhePro-----GlyAl 208
Oy 2579 GATGGCAACAGAAGGACCAACAGGCCAC-----ATCTGATAAAGGTAA 2535
Db 208 aMetGlyProArgGlyProLysGlyHisMetGlyGluArgValIleGlyHisLysGly 228
Oy 2534 GAGGGGGTG----- 2525
Db 228 uArgGlyValLysGlyLeuThrGlyProProGlyProGlyThrValIleValThrLe 248
Oy 2524 -----GATCAGCAAAAGACAGCTGCTGGGTGGGTGAGGGGACCTGCTTCTTG-- 2477
Db 248 uThrGlyProAspAsnArgThrAspLeuLysGlyGluLysGlyAspGlyAlaMetGly 268
Oy 2476 -----TGTTGCCCTCAGGACTCTTCCCTCAAAATAAGTCATATGTTCAAA 2430
Db 268 yGluProGlyProProGlyProSerGlyLeu----- 278
Oy 2429 TCCCATGGAGGAGTGTTTCATCCTAGAACTCCCATGCAAGACTACATTAACGAGCT 2370
Db 279 -ProGlyGluSer----- 282
Oy 2369 GCAGGTTAAGGGCTTAGAGATGGAAACACAGGTGACTGAGTTATTATTCAGCTCCCAAAA 2310
Db 283 -----TyrGlySerGluLysGlyAlaProGlyAsp----- 292
Oy 2309 CCCTTCTCTA-----GGTGTGCTCACTAGGAGGCTAGCTGT 2271
Db 293 -ProGlyLeuGlnGlyLysProGlyLysAspGlyValPro----- 305
Oy 2270 AACCTTGAGCCCTGGGTAAATCCACCTGCAGAGTCCCGCATTCCTAGTGGATGGAGCCCTTC 2211
Db 306 -----GlyPheProGlySerGluGlyValLysGlyLysArgGlyPheProGly 321
Oy 2210 TGGCCTCCCTGTATAGTCCACAGCTGAACACCCCTTGAAGGCCCTCCAGTCAGG-- 2156
Db 321 y-----LeuMetGlyGluAspGlyIleLysGly---GlnLysGlyAspI 335
Oy 2155 -----AGCCCTAGAGACTGGGGAGAGG----- 2132
Db 335 eGlyProGlyPheArgGlyProThrGlyTyrThrAspThrThrGlnGluLysGlyAs 355

```



```

Db 946 GluSerHisValIleGlyAspGlyGluProGlyLeuLysGlyPheAlaGlyAsn 965
QY 479 CCAGGC-----CCAGCACTGGACCAA 459
Db 966 ProGlyGlyLysGlyAsnArgGlyValProGlyMetProGlyLeuLysGlyGly 985
QY 458 TGCCAGCACCATGGTCATGAATCTCTCTCTCCCTCCACTCCAGCAGCAGAGCGGCA 399
Db 986 LeuProGlyProAlaGly-----ProGlyProArgGlyAspLeuGly 1000
QY 398 CATAGGTGATGCTGGCGCAACACACCT-----369
Db 1001 Ser-----ThrGlyAsnProGlyGluProGlyLeuArgGlyIleProGlySerMetGly 1018
QY 368 -----CCAGGCCAAAGG-----TTAGCAGGTGTACACACAGCT 333
Db 1019 AsnMetGlyMetProGlySerLysGlyLysArgGlyThrLeuGlyPheProGlyArgAla 1038
QY 332 GGCCTTCCGGTGCC-----318
Db 1039 GlyArgProGlyLeuProGlyIleHisGlyLeuGlnGlyAspLysGlyGluProGlyTyr 1058
QY 317 GCAGCAGGCGGCTCACCCACAGCCTCTGCACCATAGTGG-----CCAGGCGG 270
Db 1059 SerGluGlyThrArgProGlyProGlyProGlyProGlyProGlyLeuProGlyAs 1078
QY 269 GTAGGCTCAGGGGCGCTTCAGGCACCTCCA---GAACGTCTCTCGCTCGCTCTGCT-- 215
Db 1078 pMetGlyLysLysGlyGluMetGlyGlnProGlyProGlyHisLeuGlyProAlaGly 1098
QY 214 -CCAGAGCTGGCGCTCTCTCTCTGTCGCGCCCACTGCTAGATCAGCCAGCGC 156
Db 1098 yProGlyAlaProGlySerProGlySerProGlyLeuPro---GlyLysProGlyPr 1117
QY 155 CCATTCTCTCCACCCCTTT-----GTGCGCGTCCAGCTCTCA 117
Db 1117 oHisGlyAspLeuGlyPheLysGlyIleLysGlyLeuLeuGly-ProProGlyIleArg 1137
QY 116 GCCATGCTCACACCTGCTGCTGGGGCAC-----CTCAGTGGGGACA 72
Db 1137 lyProGlyLeuProGlyPheProGlySerProGlyProMetGlyIleArgGlyAspG 1157
QY 71 CGTCTCATCTCAGATCGTGC-----49
Db 1157 lnglyArgAspGlyIleProGlyProAlaGlyGlyGlyGlyGlyLeuLeuArg 1177
QY 48 -----CGAGGCGCGGCTGTACCCGAGCGCGCTGCAGG 10
Db 1177 laProGlyProArgGlyAsnProGlyAlaGlnGlyAlaLys 1191
RESULT 43
CALB_MOUSE
ID CALB_MOUSE STANDARD; PRT: 1804 AA.
AC Q61245; Q64047;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Collagen alpha 1(XI) chain precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=96015067; PubMed=8530046;
RA Yoshioka H., Inoguchi K., Khaleduzzaman M., Ninomiya Y.,
RA Andrikopoulos K., Ramirez F.;
RT "Coding sequence and alternative splicing of the mouse alpha 1(XI)
RL collagen gene (Col1a1).";
RL Genomics 28:337-340(1995).
RN [2]

```

```

RP SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.
RC STRAIN=C57BL/6;
RX MEDLINE=95163095; PubMed=7859283;
RA Li Y., Lacerda D.A., Warman M.L., Beller D.R., Yoshioka H.,
RA Ninomiya Y., Oxford J.T., Morris N.P., Andrikopoulos K.,
RA Ramirez F., Wardell J.B., Lifferth G.D., Teuscher C., Woodward S.R.,
RA Taylor B.A., Seemiller R.E., Olsen B.R.;
RT "A fibrillar collagen gene, Col1a1, is essential for skeletal
morphogenesis.";
RL Cell 80:423-430(1995).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
MODIFICATION OF ALPHA 1(XI). ALPHA 1(XI) CAN ALSO BE FOUND INSTEAD
OF ALPHA 3(XI)=1(XI) (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- DISEASE: DEFECTS IN COL1A1 ARE ASSOCIATED WITH CHONDRODYSPLASIA,
AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SKELETAL DEFECTS
CAUSED BY ABNORMALITIES IN THE CARTILAGE OF LIMBS, RIBS, MANDIBLES
AND TRACHEA.
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
EMBL; D38162; BAA07367.1; -;
EMBL; S74574; AAB33439.1; -;
MGD; MGI:88446; Col1a1.
InterPro; IPR000087; Collagen.
InterPro; IPR000885; Fib_collagen_C.
InterPro; IPR001791; Laminin_G.
InterPro; IPR001230; Prenyl_site.
InterPro; IPR003129; TSPN.
Pfam; PF01391; Collagen; 16.
Pfam; PF01410; COLF1; 1.
Pfam; PF02210; TSPN; 1.
ProDom; PD000007; Collagen; 1.
ProDom; PD002078; Fib_collagen_C; 1.
SMART; SM00038; COLFI; 1.
SMART; SM00282; LamG; 1.
SMART; SM00210; TSPN; 1.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Glycoprotein; Collagen; Signal; Alternative splicing;
Disease mutation.
FT SIGNAL 1 35
FT PROPEP 36 511
FT CHAIN 512 1561
FT PROPEP 1562 1804
FT DOMAIN 37 417
FT DOMAIN 418 506
FT DOMAIN 507 509
FT DOMAIN 510 527
FT DOMAIN 528 1540
FT DOMAIN 1541 1561
FT CARBOHYD 1638 1638
FT SITE 610 610
FT SITE 1450 1450
FT VARSPIC 329 413
FT VARIANT 189 195
FT VARIANT 196 1804
SQ SEQUENCE 1804 AA; 180963 MW; FE2D9DEDE1E4219A CRC64;

```

Alignment Scores:

Pred. No.: 6.91e-07 Length: 1804  
 Score: 300.50 Matches: 286  
 Percent Similarity: 27.87% Conservative: 56  
 Best Local Similarity: 23.31% Mismatches: 388  
 Query Match: 4.83% Indels: 499  
 DB: 1 Gaps: 67

US-09-759-143-110 (1-3410) x CALB\_MOUSE (1-1804)

QY	3065	CCTGAGTAGCT---CCAAACAGGTTCTGGAGCTGTGGGAAAGTTGGGGTAGGGGA	3009	1804
Db	429	ProAlaValValGluProGlyMetLeuValGlu	443	286
QY	3008	AGTTGGGGTAGGGGAATTTGGGAGTGCCTTCATCAGCCAGTCTCAGAGAGAGTA	2949	56
Db	444	ProAlaGlyProAlaGlyLeuMetGly	457	388
QY	2948	GAGGGAGTGAAGTGGGGGAACAGGCTGGCCAGAGAGAGGGTGTAGGGAG	2899	499
Db	458	GlyProSerGlyLeuProGlyAspProGlyAspArgGlyProProGlyArg	474	67
QY	2888	CGTTGAGACCTGAGCC-----CCACCTCTACCTTCTTCACACCCCT	2844	
Db	475	ProGlyLeuProGlyAlaAspGlyLeuProGlyProGlyThrMetLeuMetLeuPro	494	
QY	2843	AACCTTGGCTAACAGCATTTGCAATTTATCATTTGGGATGAGTAGAATTTCCAAGTCTCTG	2784	
Db	495	-----PheArgTyrGlyGlyAsp-----	500	
QY	2783	GGTTAGGCATTTTGGGGGCCAG-----CCCGAGAGAGAGAGATTTCTG	2739	
Db	501	-----GlySerIysGlyProThrIleSerAlaGlnGluAlaGlnAlaGln	515	
QY	2738	GCATCATCAGCCCATGACCATCTCTCAGGGAGCTGAT-----TGT	2694	
Db	516	AlaIleLeuGlnGlnAlaArgIleAlaLeuArgGlyProProGlyProMetGlyLeuThr	535	
QY	2693	TGGGATCCCCACCTCCCAATATTAGACACCAACACAGAAAGCTAGCAATGGATT	2634	
Db	536	GlyArgProGlyProValGlyGlyProGlySerThrGlyAlaLysGlyGluSerGlyAs	555	
QY	2633	CCCTTCTACTTTGTTAAATAAAGTTAAATTTAAATGCTGTCTCTCTGTATGCC	2574	
Db	555	pProGlyPro-----GlnGlyProArgGlyValGlnGln	566	
QY	2573	AACAGAGGACCAACAGGCCACATCTGATAAAGTAAAGAGGGGGTGGATCAGCAAA	2514	
Db	566	yProProGlyProThrGlyLysProGlyLysArgGlyArgProGlyAlaAspGlyArg	586	
QY	2513	AGACAGTGTGTGGGTGAGGGACCTGTTCTTGTGTGTGCTGCTCAGGACTCTTCCC	2454	
Db	586	g-----	586	
QY	2453	CTACAATAAGTCATATGTTCAATCCCATGGAGAGTGTTCATCTCAGAACTCCCAT	2394	
Db	586	-----	586	
QY	2393	GCAAGAGCTACATTAAACGAAGCTGCAGTTAAGGGCTTAGAGATGGGAAACAGGTGA	2334	
Db	587	-----GlyMetProGlyGln	591	
QY	2333	CTGAGTTTATTCAGCTCCCAAAACCCCTTCTCTAGGTGTCTCTCAACTAGGAGCTAGCT	2274	
Db	591	u-----SerGlySerLysGlyAspArgGlyPheAspGly	602	
QY	2273	GTTAAACCTGAGCTGGTAATCCATCCCTGAGAGTCCCGCATTTCCAGTGCATGGAGCCC	2214	
Db	603	-----LeuProGlyLeuProGlyAspLysGlyHisArgGlyGluArgGlyPr	618	
QY	2213	TTCTGGCTCCCTGTATAGTCCAGACTGAACCCCTTGGAGGCCCTCCAGTCA-----	2159	
Db	618	oGlnGlyProPro-----GlyLeuProGlyAspAs	628	

QY	2158	-----GGCAGCCCTAGAGACTGGGGAGA	2136
Db	628	pGlyMetArgGlyGluAspGlyGluIleGlyProArgGlyLeuProGlyGluAlaGlyPr	648
QY	2135	GAGG-----AGAGGAGCCCCAGCCAGCCAGCTGTGCAGCTACGCACCTC	2091
Db	648	oArgGlyLeuLeuGlyProArgGlyThrProGlyProProGly-----	662
QY	2090	AGCAGCAGCAGGTGGCAGCAGAGACACATTTACTTTGGCAGCAACAGAACTGCGCGCC	2031
Db	663	-----GlnProGlyIleGlyGlyIleAsp-----	670
QY	2030	AGCCCGCAGCAGCCATGGGCTAACAGGAGCGGGAGCTGGGACCC---AGTGAGCAGG	1974
Db	671	-----GlyProGlnGlyPro-----LysGlyAsnMetGlyProGlnGlyGluProGln	686
QY	1973	CCCTCCACCCCAATGTCTGGAAGTTTCTACGCTGAGTATTGGCCAAAGTCGCTCTTGT	1914
Db	686	yProPro-----	688
QY	1913	CAATACTACTCTGTAGCAAAAGTAAATGCGACACAGAC---CCAGGCTCGCGCAGAC	1858
Db	689	-----GlyGlnGlnGlyAsnProGlySerGlnGlyLe	699
QY	1857	ACCATATAGCAGTGCAGCTGCTGAGTGGGACATGAGCCATAAAGAGGATGGG	1798
Db	699	uProGlyProGlnGlyProIleGlyProProGlyGluLysGlyPro-----GlnGln	716
QY	1797	GCCACCTGGGACAGCAGAGGAGGCTATCCAGGATGCGGAGTCCAGGAGATGCCCGG	1738
Db	716	yLysProGlyLeuAlaGly-----LeuProGlyAlaAspGlyProProGlyHisProGln	734
QY	1737	CCCGAA-----CCACCTGGGCTCGGTGG-----	1713
Db	734	yLysGluGlnSerGlyGluLysGlyAlaLeuGlyProProGlyProGlnGlyProIle	754
QY	1712	-GCTCACCCACACACACA-----	1696
Db	754	eGlyTyrProGlyProArgGlyValLysGlyAlaAspGlyValArgGlyLeuLysGlySe	774
QY	1696	-----	1696
Db	774	rLysGlyGluLysGlyGluAspGlyPheProGlyPheLysGlyAspMetGlyLeuLysGln	794
QY	1695	-----CGTACGAGACATCAGAGCA-----GAGGCCCGCACA	1664
Db	794	yAspArgGlyGluValGlyGlnValGlyProArgGlyGluAspGlyProGluGlyProLys	814
QY	1663	GAGCGGGGTGGAGTGGGAGCAGCCACTGCTCCAGCAGCAGCCAGCTGCTCATTAGGAA	1604
Db	814	sglyArgAlaGlyProThrGlyAspProGlyProSerGlyGlnAlaGlyGluLysGlyLys	834
QY	1603	G---GGAGCTCCAGCTTAGG---CCTGGCAGGAAGCTGCTCATCAGGCTGCTCCTC	1553
Db	834	sLeuGlyValProGlyLeuProGlyTyrProGlyArgGln-----	847
QY	1552	ACTGCTAGCACCTCCAGTGTCCCTCGGTATTGGCAGGAGAACACCTCTCTCCCGGTG	1493
Db	848	-----GlyProLysGlySer---ThrGlyPheProGlyPheProGlyArg	861
QY	1492	GTAGAGGAGGAGCCATGTGTAGGCGAGGATCTGAGGCTGAGAGGTGAACCCCGTGA	1433
Db	861	laAsnGlyGluLysGlyAlaArg---GlyIleAlaGly-----	872
QY	1432	GGCGCTGAAGCTGTCCACAGCCAGCAGCTGTGGCAGAGGATGTGGCAGCGCAGCCAC	1373
Db	873	-----LysProGlyProArgGlyGlnArgGly-----	881
QY	1372	AGGAAAGTGCACACTGGCCAAATAGACTGCTCGAGTGCAGAAATCGCTGCACAGCGG	1313
Db	882	-----ProThrG	884
QY	1312	GTCCA-----TGACCAGAGAGACAGGAGATGGCGCACTG	1274



```

Db      884  lyProArgGlySerArgGlyAlaArgGlyProThrGlyLysProGlyProLysGlyThrs 904
QY      1273  CAGGA-----ACAGCCCGAGGCTGCCATCCGAGCGCTTCATCATAGTGTCTCCGGGC 1220
Db      904  erGlyGlyAspGlyProGlyProGlyProGlyGluArg-----Glyp 918
QY      1219  CTCGGTCCCGGCTAGCTCTGGGCACGCCCTGGTACAGCCCTCCGCCACCAAAATCCGT 1160
Db      918  roGlnGlyPro-----GlnGlyProValGlyPheProGlyProLysGlyProp 934
QY      1159  GTAACACAGCTGAAGTTCATGAGTCCATCCAGCTGCACAGCTCAGCCACGACGCGC 1100
Db      934  roGlyProAlaGlyLysAspGlyLeuProGlyHis----- 945
QY      1099  GGCAGGGTCCGGGCATCGGCACGACAGCTGTGCAGCCGGGAAGCAGGCGCCACG 1040
Db      946  -----ProGlyGlnArgGlyGluThrGlyPheGlnG 956
QY      1039  GTTCGGAAGCAAGCGGCGCGCG-----ATGCACAGCAGTGGGCGCACAGGAGG 986
Db      956  lyLysThrGlyProGlyProGlyGlyValValGlyProGlnGlyProThr-----G 974
QY      985  GSCCAGACACCCCTCTCTGGTGGTGGGCGCCAGCGCTGCTCTCAGCCACGACGAG 926
Db      974  lyGluThrGlyProIleGlyGluArgGlyHisProGly-----ProGlyProGlyG 993
QY      925  TGTGCTGCTACGACAGGTGAGCAAGATGAGGTGAGCAGGCGCAAGAGGCACTCTCT 870
Db      993  luGlnGlyLeuPro-----GlyAlaAlaGlyLysGluGlyAlaLysGlyAspProGlyp 1011
QY      869  -----CCTGGGTGCCCGCT-----AGGG 851
Db      1011  roGlnGlyLysSerGlyLysAspGlyProAlaGlyIleArgGlyPheProGlyGluArg 1031
QY      850  GG-----CCAGGGCAGCTGTGTCCATGATGCGCAGGAGGAGTACCGCCAGCCGCC 794
Db      1031  lyLeuProGlyAlaGlnGlyAlaProGlyLeuLysGlyGlyGluGlyProGlnGlyProG 1051
QY      793  AAGACTGATCATGAGGATAGACAGAGTACGCTGCGGACAGCTGGTCCGGT----- 741
Db      1051  lngly-----ProValGlySerProGlyGluArgGlySerAlaGlyAlaA 1066
QY      740  -----CCCGGAAGAGTCCAGAGCAGGCGCTCCCA-----GTGG 707
Db      1066  laGlyProIleGlyLeuProGlyArgProGlyProGlnGlyProGlyProGlyProAlaGly 1086
QY      706  AGTGAAGCACACCTGGCCACAGAGTCCAGCAGCCCGC----- 666
Db      1086  luLysGlyAlaProGlyGluLysGlyProGlnGlyProAlaGlyArgAspGlyValGlnG 1106
QY      665  -----CCAGGATGACAGTGCAGCT-----CCAGGG----- 639
Db      1106  lyProValGlyLeuProGlyProAlaGlyProAlaGlySerProGlyGluAspGlyAspL 1126
QY      638  -----GCCTGGATCCGGGCACAGCCCTGCTAGCCAGCCCG----- 600
Db      1126  ysGlyGluIleGlyGluProGlyGlnLysGlySerLysGlyAspLysGlyGluAsnGlyp 1146
QY      599  -----CCCTTGGATGAGAAGAGGCTCAGCAGGATCCCGCAGGACAGTCCCGCAT 548
Db      1146  roProGlyProGly-----LeuGlnGlyProValGlyAlaProGlyIle----- 1161
QY      547  GAAGGGCGCGCGGCGCATAGCGTCCACCCAGTGGTCTACTGGCTGAGCCCTAGGA----- 492
Db      1162  -----AlaGlyGlyAspGlyGluProGlyProArgGlyGlnGlnGlyMetPheGlyGlnL 1180
QY      491  -----GGGGCACACAGA-----CCAGGCCCA-----GCAGTGGACCAAT 458
Db      1180  ysGlyAspGluGlyAlaArgGlyPheProGlyLeuProGlyProIleGlyLeuGlnGlyL 1200
QY      457  GCCCAGCACCATGGTGCATCACTCT-----CCTC 428

```

```

1200 euProGlyProProGlyGluLysGlyGluAsnGlyAspValGlyProMetGlyProProG 1220
427 TACCCCTTCCTCCAGCAGCAGAGCGGCACATAGTGTATGCTCGCGCCCAACACACTC 368
1220 lyProProGlyProArgGlyProGlnGlyProAsnGlyAlaAspGlyProGlnGlyProp 1240
367 CAGGCCAA-----AGGTTAGCAGGTTGACCAAGAGAGCTGGGCTTTCCGGTG 320
1240 roGlySerIleLysValGlyValGlyValGly-AspLysGlyGlu----- 1254
319 CCGCAGCAGCGGCTCACCACAGCTCTGACCATAGTGGGCCAGCGGG----- 269
1255 ProGlyGluAlaGlyAsnProGlyProProGlyGluAlaGlySerGlyLeuLysGly 1274
268 ---TAGGCTCAGGGGCGGTTCCAGGCACCTCCAGAACTGCTTCTCTCGGCTCTGCTCCA 212
1275 GluArgGlyGluLysGlyGluAlaGlyProPro-----GlyAlaAlaGly 1289
211 GAAGCTCGGGCTCTCTCTCTGCTGCTGCGCCCACTGCTAGGAATCAGCCAGGCGCCAT 152
1290 ProAlaGlyLysGlyProProGlyAspGlyProLysGlyAsnProGly----- 1307
151 TTCTGCCAGCGCTTTGGT-----GCCGCTCCAGCTTCTCAGCCCATGCTCAAC 104
1308 -----ProValGlyPheProGlyAspProGlyProProGlyGluPro-----Gly 1322
103 ACCTGCTGCTGCTGGGCGCACCTCAGTGGGACACCTCCTCATCAGTCTGAGTCTGGCCGA-- 46
1323 -ProAlaGlyGlnAspGlyValGlyGlyAspLysGlyGluAspGlyAspProGlyGlnPr 1342
45 -----GCCGCGCGGCTGTC 32
1342 oGlyProProGlyProSerGlyGluAlaGlyProProGlyProProGlyLysArgGlyPr 1362
31 ACCCGGAGCGCAGC 19
1362 oProGlyAlaSer 1366

```

RESULT 44  
CALA\_MOUSE  
ID CALA\_MOUSE STANDARD; PRT; 680 AA.  
AC Q05306;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DE 01-NOV-1995 (Rel. 32, Last annotation update)  
GN COL10A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c;  
RX MEDLINE=93143676; PubMed=8424763;  
RA Elima K., Berola I., Rosati R., Metsaranta M., Garofalo S., Perala M.,  
RA de Crombrughe B., Vuorio E.;  
RT "The mouse collagen X gene: complete nucleotide sequence, exon  
structure and expression pattern.";  
RL Biochem. J. 289:247-253(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV; TISSUE=Liver;  
RX MEDLINE=93238750; PubMed=8477738;  
RA Kong R.Y.C., Kwan K.M., Lau E.T., Thomas J.T., Boot-Handford R.P.,  
RA Grant M.E., Cheah K.S.E.;  
RT "Intron-exon structure, alternative use of promoter and expression of  
the mouse collagen X gene, Col10a-1.";  
RL Eur. J. Biochem. 213:99-111(1993).  
RN [3]  
RP SEQUENCE OF 51-680 FROM N.A.  
RC STRAIN=DBA/2J;  
RX MEDLINE=92267014; PubMed=1587271;





Db 299 -GlyLysGlnGlyLeuProGly---LeuArgGlyGlnArgGlyProAlaGlyLeuProGly 317  
 QY 1191 GCCTGCCAGAGTGCAGCGGCGCCAGGCGCCGAGACACATATGATGAAGCGTTCGGA 1250  
 Db 317 y---AlaProGlyAlaLysGlyGluArgGlyProAla 328  
 QY 1251 TGGGCGAGCTGGGGCTGTTCTGCAGTGCAGCGCCATCCCTGGCTTCTCTGTCGTCATGG 1310  
 Db 329 -GlyHisProGlyGlu---ProGlyLeuProGlySerProGlyAsnMetGlyProGlnGly 347  
 QY 1311 ACCGGCTGTCGAGCAGTTCGGCAC-----TCGACGAGTCTATT 1349  
 Db 347 yProLysGlyLysProGlyAsnHisGlyLeuProGlyAlaLysGlyGluLeuGlyLeuVa 367  
 QY 1350 TGGCAGTGTGGCAGCTTCCCTGTGGCTGGCGGCGCCACATGCTGTCACACAGTGTGG 1409  
 Db 367 LGlyProAlaGly-----ProGlyAlaArgGlyAlaArgGlyProPro-----Gly 383  
 QY 1410 CCGTGGTGACAGCTTCAGCGGCTCAGCGGCTTCCCTCTCAGCCCT----- 1458  
 Db 383 yLeuAspGlyLysThrGlyTyrProGlyGluProGlyLeuAsnGlyProLysGlyAsnPr 403  
 QY 1459 -----GCAGATCCCTGCCCTA----- 1473  
 Db 403 oGlyLeuProGlyGlnLysGlyAspProGlyValGlyGlyThrProGlyLeuArgGlyPr 423  
 QY 1474 -----CACACTGGCTCCCTCTACACCGGG 1499  
 Db 423 oValGlyProValGlyAlaLysGlyValProGlyHisAsnGlyGluAlaGlyProArgGly 443  
 QY 1500 AGAAGCAGGTGTTCTGCGCCAAATACCAGGGGACACTGGA----- 1540  
 Db 443 yGluProGlyLysProGly-Thr---ArgGlyProThrGlyProGlyValProGlyP 462  
 QY 1541 -----GGTGCTAGCAGTGAGGACAGCAGCTGTAT 1568  
 Db 462 heProGlySerLysGlyAspProGlyAsnProGlyAlaProGlyProAlaGlyLysAla 482  
 QY 1569 CGAGCTTCTGCGAGGCGCT-----AAGCTGAGGCTCCCTTCCCTAATGGACACCTGG 1622  
 Db 482 hrLysGlyLeuAsnGlyProThrGlyProGlyProGlyProGlyProGlyHisSerG 502  
 QY 1623 GTGCTGGAGGAGTGGCTGCTCCACCTCCACCGGCTGCTGGGGGCTCTCCCTGTG 1682  
 Db 502 yGluProGly-----LeuProGlyProPro----- 510  
 QY 1683 ATGCTCGTACGTGTGTGTGTGGTGAGCGCCAGCGGCTGTTCCGGGCGGG 1742  
 Db 511 -----GlyProGlyPro-----ProGlyGln- 518  
 QY 1743 GCATCTGCTGGAGCTGCCATCTCGATAGTGCCTTCTGCTGCCAGGTGGCCCAT 1802  
 Db 519 -----AlaValMetProAspGlyPheLeuLysAlaGlyGlnArgProA 533  
 QY 1803 CCGCTGTTATGGCTCATGTGTCAGCTCAGCAGCTGTCACCTGCTATATGTTGCTG 1862  
 Db 533 rgLeuSerGlyMetProLeuValSerAlaAsnHisGlyValThrGlyMetProValSerA 553  
 QY 1863 CCGCAGCGCTGGTGTGGTCCCATTTACTTGTCT-----ACACAGGTAGTATTGACA 1916  
 Db 553 laPheThrValLeuSerLysAlaTyrProAlaValGlyAlaProIleProPheAspG 573  
 QY 1917 AG 1918  
 Db 573 lu 573

RESULT 45  
 CALH\_HUMAN  
 ID CALH\_HUMAN STANDARD; PRT; 1516 AA.  
 AC P39060; Q9Y608; Q9Y607; Q9UK38;  
 DT 01-FEB-1995 (Rel. 31, Created).  
 DT 16-OCT-2001 (Rel. 40, Last sequence update).  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE GN Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].  
 OS COL18A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98164096; PubMed=9503365;  
 RA Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;  
 RT "Complete primary structure of two variant forms of human type XVIII  
 collagen and tissue-specific differences in the expression of the  
 corresponding transcripts";  
 RL Matrix Biol. 16:319-328(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2089799; PubMed=10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
 Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
 Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,  
 Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
 Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
 Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
 Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
 Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,  
 Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
 Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,  
 Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
 Lehraich H., Reinhardt R., Yaspo M.-L.;  
 RA "The DNA sequence of human chromosome 21";  
 RL Nature 405:311-319(2000).  
 [3]  
 RP SEQUENCE OF 834-1516 FROM N.A.  
 RX MEDLINE=94245237; PubMed=8188291;  
 RA Oh S.-P., Warman M.L., Seidman M.F., Cheng S., Knoll J.H., Timmons S.,  
 Olson B.R.;  
 RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen  
 and localization of the alpha 1(XVIII) collagen gene to mouse  
 chromosome 10 and human chromosome 21";  
 RL Genomics 19:494-499(1994).  
 [4]  
 RP SEQUENCE OF 1334-1516 FROM N.A.  
 RX TISSUE=Placenta;  
 RA Zhi-Yong H., Bao L., Wei-Jie Z., Xiang-Fu W.;  
 RT "Cloning and expression of human endostatin gene in Escherichia  
 coli";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP INVOLVEMENT IN KNOBLOCH SYNDROME.  
 RX MEDLINE=20400145; PubMed=10942434;  
 RA Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,  
 Passos-Bueno M.R.;  
 RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis  
 and tumor growth, plays a critical role in the maintenance of retinal  
 structure and in neural tube closure";  
 RL Hum. Mol. Genet. 9:2051-2058(2000).  
 [6]  
 RP VARIANT ASN-1437.  
 RX MEDLINE=21518361; PubMed=11606364;  
 RA Lugheetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,  
 Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,  
 Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;  
 RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes  
 for the development of prostatic adenocarcinoma";  
 RL Cancer Res. 61:7375-7378(2001).  
 CC -!- FUNCTION: COL18A1 PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE  
 RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.  
 CC -!- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL  
 PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
 BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH  
 FACTOR SIGNALLING.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM/NCI-303 AND A LONG  
 FORM/NC-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.







Db 2221 ATGCACGTGAATGCGGCACTCTGCAGGTGATACCCAGGCTCAGGGTTACAGCTAGC 2280  
Qy 2281 CTCCTACTCTGAGACACACTAGAGAGGGTTTTGGAGCTGAATAAATCAGTACCTG 2340  
Db 2281 CTCCTAGTCTGAGACACACTAGAGAGGGTTTTGGAGCTGAATAAATCAGTACCTG 2340  
Qy 2341 GTTTCCTACTCTAAGCCCTTAACCTGACGCTTCGTTAATGATGCTCTGCATGGGAG 2400  
Db 2341 GTTTCCTACTCTAAGCCCTTAACCTGACGCTTCGTTAATGATGCTCTGCATGGGAG 2400  
Qy 2401 TTTCTAGGATGAACACTCTCTCCATGGGATTTGAACATATGACTTATTTGTAGGGGAAGA 2460  
Db 2401 TTTCTAGGATGAACACTCTCTCCATGGGATTTGAACATATGACTTATTTGTAGGGGAAGA 2460  
Qy 2461 GTCTCTAGGGGGAACACACAAAGAACAGGTCCCTCAGCCACAGCAGCTGTCTTTTGGT 2520  
Db 2461 GTCTCTAGGGGGAACACACAAAGAACAGGTCCCTCAGCCACAGCAGCTGTCTTTTGGT 2520  
Qy 2521 GATCCACCCCTCTTACCTTTATCAGGATGTGGCTGTGTGCTCTCTGTGTCATCA 2580  
Db 2521 GATCCACCCCTCTTACCTTTATCAGGATGTGGCTGTGTGCTCTCTGTGTCATCA 2580  
Qy 2581 CAGACACACAGCATTAATATTTAACTTATTTTAACTTATTTTAACTTATTTTAACTTATTTT 2640  
Db 2581 CAGACACACAGCATTAATATTTAACTTATTTTAACTTATTTTAACTTATTTTAACTTATTTT 2640  
Qy 2641 TGTAGCTTTTCTGTGCTGTCTTAATATTTGGGTAGGTGGGGATCCCAACAATCA 2700  
Db 2641 TGTAGCTTTTCTGTGCTGTCTTAATATTTGGGTAGGTGGGGATCCCAACAATCA 2700  
Qy 2701 GGTCCCTCAGATAGTGTGCTATTTGGGTGATCATTTGCCAGATCTTCTCTCTGGGGT 2760  
Db 2701 GGTCCCTCAGATAGTGTGCTATTTGGGTGATCATTTGCCAGATCTTCTCTCTGGGGT 2760  
Qy 2761 CTGGCCCTCCCAATAGCTTAACCCAGGACCTTGGAAATCTACTCATCCCAATGATAAT 2820  
Db 2761 CTGGCCCTCCCAATAGCTTAACCCAGGACCTTGGAAATCTACTCATCCCAATGATAAT 2820  
Qy 2821 TCCAAATGTGTTACCCAGGTTAGGTTGTTGAAGGAAGTGAAGGTTGGGCTTCAGGT 2880  
Db 2821 TCCAAATGTGTTACCCAGGTTAGGTTGTTGAAGGAAGTGAAGGTTGGGCTTCAGGT 2880  
Qy 2881 CTCACAGGCTTCCCTTAACCCACCTTCTCTCTGTGGCCAGGCTGTTTCCCTCCCTTCCA 2940  
Db 2881 CTCACAGGCTTCCCTTAACCCACCTTCTCTCTGTGGCCAGGCTGTTTCCCTCCCTTCCA 2940  
Qy 2941 CTCCTCTACTCTCTTAGGACTGGGCTGATGAAGGACCTGCCCAAAATTTCCCTACC 3000  
Db 2941 CTCCTCTACTCTCTTAGGACTGGGCTGATGAAGGACCTGCCCAAAATTTCCCTACC 3000  
Qy 3001 CCCAACTTCCCTTACCCCAACTTTCCCAACAGCTCCCAACCTGTTTGGAGCTACT 3060  
Db 3001 CCCAACTTCCCTTACCCCAACTTTCCCAACAGCTCCCAACCTGTTTGGAGCTACT 3060  
Qy 3061 GCAGGACAGAGACAAAGTGGGTTTCCCAAGCCTTTGTCATCTCAGCCCCCAGT 3120  
Db 3061 GCAGGACAGAGACAAAGTGGGTTTCCCAAGCCTTTGTCATCTCAGCCCCCAGT 3120  
Qy 3121 ATATCTGTGCTGGGGAATCTCACACAGAACTCAGGAGCACCCCTGCTGAGCTAAGG 3180  
Db 3121 ATATCTGTGCTGGGGAATCTCACACAGAACTCAGGAGCACCCCTGCTGAGCTAAGG 3180  
Qy 3181 GAGGCTTATCTCTCAGGGGGGTTTAAAGTCCCTTTGCAATATATGCTTATTTATT 3240  
Db 3181 GAGGCTTATCTCTCAGGGGGGTTTAAAGTCCCTTTGCAATATATGCTTATTTATT 3240  
Qy 3241 TAGGGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGTATATGTTTATGGTACA 3300  
Db 3241 TAGGGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGTATATGTTTATGGTACA 3300  
Qy 3301 AAATTAAGGCTTTCTTATATGTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 3360  
Db 3301 AAATTAAGGCTTTCTTATATGTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 3360

Db 3301 AAATTAAGGCTTTCTTATATGTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 3360  
Qy 3361 AA 3410  
Db 3361 AA 3410

## RESULT 2

US-09-030-607-110  
; Sequence 110, Application US/09030607  
; Patent No. 6262245  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS  
; NUMBER OF SEQUENCES: 224  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/030,607  
; FILING DATE: 25-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Makl, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.427C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 110:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3410 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-09-030-607-110

Query Match 100.0%; Score 3409.6; DB 4; Length 3410;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGAACACAGCTCTCAGCGCTGCTCCGGGTGACAGCCGCGCCCTCGCCAGGATCTCA 60  
Db 1 GGGAACACAGCTCTCAGCGCTGCTCCGGGTGACAGCCGCGCCCTCGCCAGGATCTCA 60  
Qy 61 GTGATGAGAGCTGTCTCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGGCTGAG 120  
Db 61 GTGATGAGAGCTGTCTCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGGCTGAG 120  
Qy 121 AAGCTGGACCGGACCAAGGCTGCAGAAATGGCGCTTGGCTGATTCCTAGGAGTT 180  
Db 121 AAGCTGGACCGGACCAAGGCTGCAGAAATGGCGCTTGGCTGATTCCTAGGAGTT 180  
Qy 181 GGCGGACCAAGGAGGAGGCGCGAGCTTCTGGAGCAGAGCCGAGCAAGCACTTCTG 240  
Db 181 GGCGGACCAAGGAGGAGGCGCGAGCTTCTGGAGCAGAGCCGAGCAAGCACTTCTG 240  
Qy 241 GAGTGCCTCAACGGGCCCTTGAGCCCTACCCGCCCTGGCCCACTATGGTCCAGAGCTGTG 300  
Db 241 GAGTGCCTCAACGGGCCCTTGAGCCCTACCCGCCCTGGCCCACTATGGTCCAGAGCTGTG 300







QY	661	CGTGGCGTGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCT	720	QY	1741	GGCATCTGCTGGACCTCGCCATCTCGATAGTGCCTTCTGCTGCTCCAGGTGGCCCC	1800
DB	661	CGTGGCGTGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCT	720	DB	1741	GGCATCTGCTGGACCTCGCCATCTCGATAGTGCCTTCTGCTGCTCCAGGTGGCCCC	1800
QY	721	GCTCTCTGACCTCTTCCGGGACCGGACCACTGTGCGCAGGCTTACTGTCTATGCTTT	780	QY	1801	ATCCCTGTTTATGGGCTCCATTTCTCAGCTCAGCAGTCTGTCTACCTGCTATATGCTGTC	1860
DB	721	GCTCTCTGACCTCTTCCGGGACCGGACCACTGTGCGCAGGCTTACTGTCTATGCTTT	780	DB	1801	ATCCCTGTTTATGGGCTCCATTTCTCAGCTCAGCAGTCTGTCTACCTGCTATATGCTGTC	1860
QY	781	CATGATCAGTCTTGGGGCTGCCTGGGTACTCTCTGCTGCCATTTGACTGGGACACAG	840	QY	1861	TGCGGAGGCTGGGTCTGGTGGCCATTTACTTTGTCTACACAGGTAGTATTTTGACAAGAG	1920
DB	781	CATGATCAGTCTTGGGGCTGCCTGGGTACTCTCTGCTGCCATTTGACTGGGACACAG	840	DB	1861	TGCGGAGGCTGGGTCTGGTGGCCATTTACTTTGTCTACACAGGTAGTATTTTGACAAGAG	1920
QY	841	TGCGCTGGCCCTTACTTGGGACCCAGGAGGAGTGCCTCTTTGGCTGTGCTCAGCCCTAT	900	QY	1921	CGACTTTGGCCAAATACTCAGCTAGAAAACCTCCAGACATTTGGGTGGAGGSCCTGCT	1980
DB	841	TGCGCTGGCCCTTACTTGGGACCCAGGAGGAGTGCCTCTTTGGCTGTGCTCAGCCCTAT	900	DB	1921	CGACTTTGGCCAAATACTCAGCTAGAAAACCTCCAGACATTTGGGTGGAGGSCCTGCT	1980
QY	901	CTTCTCTACCTCGGTAGCAGCCACACTGCTGGTGGCTGAGGAGGAGCGCTGGGCCCCAC	960	QY	1981	CAGTGGTCCAGCTCCCGCTCTCTGTAGCCCATGGGCTGCCGGCTGGCGGCCAGT	2040
DB	901	CTTCTCTACCTCGGTAGCAGCCACACTGCTGGTGGCTGAGGAGGAGCGCTGGGCCCCAC	960	DB	1981	CAGTGGTCCAGCTCCCGCTCTCTGTAGCCCATGGGCTGCCGGCTGGCGGCCAGT	2040
QY	961	CGAGCCAGCAGAGGGCTGTGGGCCCCCTCTTGTGCCCCACTGCTGTCTATGCCGGGC	1020	QY	2041	TTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGTGCTGCCACCTGCTGCTGAGTGGCTA	2100
DB	961	CGAGCCAGCAGAGGGCTGTGGGCCCCCTCTTGTGCCCCACTGCTGTCTATGCCGGGC	1020	DB	2041	TTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGTGCTGCCACCTGCTGCTGAGTGGCTA	2100
QY	1021	CGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCGGCTGCAACAGCTGTGCTGCCG	1080	QY	2101	GCTGCACAGCTGGGGCTGGGGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2160
DB	1021	CGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCGGCTGCAACAGCTGTGCTGCCG	1080	DB	2101	GCTGCACAGCTGGGGCTGGGGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2160
QY	1081	CATGCCCGCAGCCTGGCGGCTCTGTGGCTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCT	1140	QY	2161	ACTGAGGCTTCCAAAGGGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2220
DB	1081	CATGCCCGCAGCCTGGCGGCTCTGTGGCTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCT	1140	DB	2161	ACTGAGGCTTCCAAAGGGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2220
QY	1141	GACCTTCAGCTGTTTACACGATTTCTGCGGAGGGCTGTACACAGGCTGCTGCCAG	1200	QY	2221	ATGCACTGGAATCGGGGACTCTCAGGTGGATTTACCCAGGCTCAGGGTTTAAACAGTAGC	2280
DB	1141	GACCTTCAGCTGTTTACACGATTTCTGCGGAGGGCTGTACACAGGCTGCTGCCAG	1200	DB	2221	ATGCACTGGAATCGGGGACTCTCAGGTGGATTTACCCAGGCTCAGGGTTTAAACAGTAGC	2280
QY	1201	AGCTGAGCGGGCAGCGAGCCCGGAGACACTATGATGAAGCGCTTCGATGGGAGGCT	1260	QY	2281	CTCTAGTTGAGACACACTTAGAGAAGGTTTGGGAGCTGAATAAATCACTCAGTCACTG	2340
DB	1201	AGCTGAGCGGGCAGCGAGCCCGGAGACACTATGATGAAGCGCTTCGATGGGAGGCT	1260	DB	2281	CTCTAGTTGAGACACACTTAGAGAAGGTTTGGGAGCTGAATAAATCACTCAGTCACTG	2340
QY	1261	GGGCTGTTCTGCTGAGTGGCCATCTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT	1320	QY	2341	GTTCCTCATCTTAAGCCCTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2400
DB	1261	GGGCTGTTCTGCTGAGTGGCCATCTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT	1320	DB	2341	GTTCCTCATCTTAAGCCCTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2400
QY	1321	GCAGCATTCGGCACTCGACAGTCTATTGGCCAGTGTGGAGCTTTCCCTGCTGCTGCT	1380	QY	2401	TTTCTAGATGAACACTCTCCATGGATTTGAACATATGACTTATTTTGTAGGGGAAGA	2460
DB	1321	GCAGCATTCGGCACTCGACAGTCTATTGGCCAGTGTGGAGCTTTCCCTGCTGCTGCT	1380	DB	2401	TTTCTAGATGAACACTCTCCATGGATTTGAACATATGACTTATTTTGTAGGGGAAGA	2460
QY	1381	CGGTGCCACATGCTGTCCACAGTGTGGCCGTGTGACAGCTTACAGCCGCTCAGCGG	1440	QY	2461	GTCTGAGGGGCAACACAAAGAACCCAGGTCCCTCAGCCCAACAGCACTGTCTTTTGTCT	2520
DB	1381	CGGTGCCACATGCTGTCCACAGTGTGGCCGTGTGACAGCTTACAGCCGCTCAGCGG	1440	DB	2461	GTCTGAGGGGCAACACAAAGAACCCAGGTCCCTCAGCCCAACAGCACTGTCTTTTGTCT	2520
QY	1441	GTTTCACTTCTGAGCCCTGAGATCTTCCCTACACTGGCTCCCTCTTACACCGGGA	1500	QY	2521	GATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTTGGTCTCTCTCTCTCTCTCTCT	2580
DB	1441	GTTTCACTTCTGAGCCCTGAGATCTTCCCTACACTGGCTCCCTCTTACACCGGGA	1500	DB	2521	GATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTTGGTCTCTCTCTCTCTCTCTCT	2580
QY	1501	GAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGGAGCAG	1560	QY	2581	CAGACACAGCAGCTTTAAATATTTAACTTATTTTAAACAAAGTAGAAGGGGAATCCAT	2640
DB	1501	GAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGGAGCAG	1560	DB	2581	CAGACACAGCAGCTTTAAATATTTAACTTATTTTAAACAAAGTAGAAGGGGAATCCAT	2640
QY	1561	CCTGATGACAGCTTCTGCGAGCCCTTACGCTTGGAGTCCCTTCCCTTAAATGGACAGCT	1620	QY	2641	TGCTAGCTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2700
DB	1561	CCTGATGACAGCTTCTGCGAGCCCTTACGCTTGGAGTCCCTTCCCTTAAATGGACAGCT	1620	DB	2641	TGCTAGCTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2700
QY	1621	GGGTGCTGGAGGAGTGGCTGCTTCCACCTCCACCGGCTCTGCGGGGCTCTGCTGCTG	1680	QY	2701	GTCCTCCCTGAGTACCTGGTCAATTTGGGCTGATCTTGGCAGAACTTCTCTCTCTCTCTCT	2760
DB	1621	GGGTGCTGGAGGAGTGGCTGCTTCCACCTCCACCGGCTCTGCGGGGCTCTGCTGCTG	1680	DB	2701	GTCCTCCCTGAGTACCTGGTCAATTTGGGCTGATCTTGGCAGAACTTCTCTCTCTCTCTCT	2760
QY	1681	TGATGCTCTCCGTACGTGTGTGGGTGAGCCACCGAGGAGGCTGTTCCCGGGCGG	1740	QY	2761	CTGGCCCCCAAAATGCCTTAACCCAGGACCTTGGAAATTTCTACTCATCCCCAAATGATAAT	2820
DB	1681	TGATGCTCTCCGTACGTGTGTGGGTGAGCCACCGAGGAGGCTGTTCCCGGGCGG	1740	DB	2761	CTGGCCCCCAAAATGCCTTAACCCAGGACCTTGGAAATTTCTACTCATCCCCAAATGATAAT	2820
				QY	2821	TCCAAATGCTGTTCACCAAGGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGGCTTCAGGT	2880

Db 2821 TCCAAATGCTGTACCAAGGTTAGGGTGTGGAAGAGGTAGAGGGTGGGGCTCAGGT 2880  
QY 2881 CTCACGGCTTCCCTAACACCCCTCTTCTCTTGGGCCAGCGTGGTCCCGCCACTTCCA 2940  
Db 2881 CTCACGGCTTCCCTAACACCCCTCTTCTCTTGGGCCAGCGTGGTCCCGCCACTTCCA 2940  
QY 2941 CTCCTCTCTACTCTCTAGGACTGGGCTGATGAAGGCACTGGCCCAAAATTTCCCTTACC 3000  
Db 2941 CTCCTCTCTACTCTCTAGGACTGGGCTGATGAAGGCACTGGCCCAAAATTTCCCTTACC 3000  
QY 3001 CCCACTTTCCTTACCCCTAACCTTCCCAACAGCTCTCCCAACCCCTGTTTGGAGCTACT 3060  
Db 3001 CCCACTTTCCTTACCCCTAACCTTCCCAACAGCTCTCCCAACCCCTGTTTGGAGCTACT 3060  
QY 3061 GCAGACACAGAGCAAAAGTGGGCTTCCCAAGCTTTCCTCATCTCAGCCCCCAGAGT 3120  
Db 3061 GCAGACACAGAGCAAAAGTGGGCTTCCCAAGCTTTCCTCATCTCAGCCCCCAGAGT 3120  
QY 3121 ATATCTGTGCTGGGGAATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGG 3180  
Db 3121 ATATCTGTGCTGGGGAATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGG 3180  
QY 3181 GAGGCTTATCTCTCAGGGGGGTTAAGTCCCGTTTGCATTAATGTCGTTATTATT 3240  
Db 3181 GAGGCTTATCTCTCAGGGGGGTTAAGTCCCGTTTGCATTAATGTCGTTATTATT 3240  
QY 3241 TAGCGGGTGAATATTTATATCTTAAGTGAGCAATCAGACTATAATGTTTATGGTGACA 3300  
Db 3241 TAGCGGGTGAATATTTATATCTTAAGTGAGCAATCAGACTATAATGTTTATGGTGACA 3300  
QY 3301 AAATAAAGGCTTCTTATATGTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 3360  
Db 3301 AAATAAAGGCTTCTTATATGTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 3360  
QY 3361 AA 3410  
Db 3361 AA 3410

## RESULT 4

US-09-439-313-110

; Sequence 110, Application US/09439313

; Patent No. 6329505

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan Louise

; APPLICANT: Jiang Yugu

; APPLICANT: Reed, Steven G.

; APPLICANT: Kalos, Michael

; APPLICANT: Fanger, Gary

; APPLICANT: Retter, Mark

; APPLICANT: Solk, John

; APPLICANT: Day, Craig

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427C9

; CURRENT APPLICATION NUMBER: US/09/439,313

; CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 575

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 110

; LENGTH: 3410

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-439-313-110

Query Match

Best Local Similarity 100.0%; Score 3409.6; DB 4; Length 3410;

Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAACACAGCTGCACGCGCTGCGGCTGAGAGCCGCGCGCTCGGCGAGGATCTGA 60  
Db 1 GGAACACAGCTGCACGCGCTGCGGCTGAGAGCCGCGCGCTCGGCGAGGATCTGA 60  
QY 61 GTGATGAGAGCTGCCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGGTGAG 120  
Db 61 GTGATGAGAGCTGCCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGGTGAG 120  
QY 121 AAGTGTACCGCGGACCAAAAGGCTGGCAGAAATGGCGCTGCTGATTCCTAGGAGTT 180  
Db 121 AAGTGTACCGCGGACCAAAAGGCTGGCAGAAATGGCGCTGCTGATTCCTAGGAGTT 180  
QY 181 GGGGAGCAGAGGAGGAGGCGCAGCTTCTGGAGCAGAGCGAGCAGCAAGCAGTCTTG 240  
Db 181 GGGGAGCAGAGGAGGAGGCGCAGCTTCTGGAGCAGAGCGAGCAGCAAGCAGTCTTG 240  
QY 241 GAGTGTCTGAACGCGCCCTGAGCCCTACCCGCTGCCCCACTATGGTTCAGAGGCTG 300  
Db 241 GAGTGTCTGAACGCGCCCTGAGCCCTACCCGCTGCCCCACTATGGTTCAGAGGCTG 300  
QY 301 GGTGAGCGCGCTGCTGCGGACCGGAAAGCCAGCTTCTGGTGCACCTGCTAACCTT 360  
Db 301 GGTGAGCGCGCTGCTGCGGACCGGAAAGCCAGCTTCTGGTGCACCTGCTAACCTT 360  
QY 361 TGGCCTGGAGGTGTTTGGCCAGGCAATCAGCTATGTGCGCGCTCTGCTGCTGGAAGT 420  
Db 361 TGGCCTGGAGGTGTTTGGCCAGGCAATCAGCTATGTGCGCGCTCTGCTGCTGGAAGT 420  
QY 421 GGGGTAGAGGAGAGTTCATGACCATGCTGGGCAATGTTGCCATGTTGCCAGTGT 480  
Db 421 GGGGTAGAGGAGAGTTCATGACCATGCTGGGCAATGTTGCCATGTTGCCAGTGT 480  
QY 481 CTGTGTCGCGCTCTAGGCTCAGCCAGTGCACCTGCGGTGAGCAGCTATGGCGCGCG 540  
Db 481 CTGTGTCGCGCTCTAGGCTCAGCCAGTGCACCTGCGGTGAGCAGCTATGGCGCGCG 540  
QY 541 GCCTTTCATCTGGGCACTGCTGTTGGCAATGCTGTTGCCATGTTGCCAGTGT 600  
Db 541 GCCTTTCATCTGGGCACTGCTGTTGGCAATGCTGTTGCCATGTTGCCAGTGT 600  
QY 601 CGGCTGGCTAGCAGGCTGCTGTCGCGGACCGGACCTCCAGGCGCTGAGCTGGCACTGCTCAT 660  
Db 601 CGGCTGGCTAGCAGGCTGCTGTCGCGGACCGGACCTCCAGGCGCTGAGCTGGCACTGCTCAT 660  
QY 661 CCTGGCGCTGGGCTGCTGAGCTTCTGTGGCAGAGTGTGCTTCTACCTCCAGGCGCT 720  
Db 661 CCTGGCGCTGGGCTGCTGAGCTTCTGTGGCAGAGTGTGCTTCTACCTCCAGGCGCT 720  
QY 721 GCTCTCTGACCTCTTCCGCGGACCGGACCTGCTGCGGAGCTGCTGCTATGCGCTT 780  
Db 721 GCTCTCTGACCTCTTCCGCGGACCGGACCTGCTGCGGAGCTGCTGCTATGCGCTT 780  
QY 781 CATGATCAGTCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
Db 781 CATGATCAGTCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
QY 841 TGCCCTGCGCGCTTACCTGGGCACTGAGGAGTGTGCTTGGCGCTGCTGCTGCTGCTGCT 900  
Db 841 TGCCCTGCGCGCTTACCTGGGCACTGAGGAGTGTGCTTGGCGCTGCTGCTGCTGCTGCT 900  
QY 901 CTTCCTACCTGCTGAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Db 901 CTTCCTACCTGCTGAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
QY 961 CGAGCCAGCAGAGGCTGCTGCGGCCCTCTCTGTCGCGCCCTGCTGCTGCTGCTGCTGCTGCT 1020  
Db 961 CGAGCCAGCAGAGGCTGCTGCGGCCCTCTCTGTCGCGCCCTGCTGCTGCTGCTGCTGCTGCT 1020  
QY 1021 CCGCTTGGCTTTCGGAACCTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
Db 1021 CCGCTTGGCTTTCGGAACCTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
QY 1081 CATGCCCGCACCTTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140













QY	2221	ATGCACTGGAATGCGGGGACTCTCGAGGTGATTAACCCAGGCTCAGGGTTAAACAGCTAGC	2280
Db	2221	ATGCACTGGAATGCGGGGACTCTCGAGGTGATTAACCCAGGCTCAGGGTTAAACAGCTAGC	2280
QY	2281	CTCCTAGTTGAGACACACTAGAGAAGGTTTTTGGGAGCTGAATAAACTCAGTCACCTG	2340
Db	2281	CTCCTAGTTGAGACACACTAGAGAAGGTTTTTGGGAGCTGAATAAACTCAGTCACCTG	2340
QY	2341	GTTCCTCCATCTCTAAGCCCTTAAACCTGCAGCTTCGTTTTAATGTAGCTCTTGATGGGAG	2400
Db	2341	GTTCCTCCATCTCTAAGCCCTTAAACCTGCAGCTTCGTTTTAATGTAGCTCTTGATGGGAG	2400
QY	2401	TTTCTAGGATGAACACACTCCCTCATGGGATTTGAACATATGACTATTGTAGGGGAAGA	2460
Db	2401	TTTCTAGGATGAACACACTCCCTCATGGGATTTGAACATATGACTATTGTAGGGGAAGA	2460
QY	2461	GTCTGAGGGGAACACACAGAACCAGGTCCCTTCAGCCACAGCACTGTCTTTTTGCT	2520
Db	2461	GTCTGAGGGGAACACACAGAACCAGGTCCCTTCAGCCACAGCACTGTCTTTTTGCT	2520
QY	2521	GATCCACCCCTCTTACCTTTATCAGATGTGCGCTGTGGTCTCTCTGTCGCCATCA	2580
Db	2521	GATCCACCCCTCTTACCTTTATCAGATGTGCGCTGTGGTCTCTCTGTCGCCATCA	2580
QY	2581	CAGAGACACAGGCATTAAATATTAACTTATTATTTAACAAGTAGAAGGAATCCAT	2640
Db	2581	CAGAGACACAGGCATTAAATATTAACTTATTATTTAACAAGTAGAAGGAATCCAT	2640
QY	2641	TGCTAGCTTTCTGTGTTGGTCTTAATATTGGCTAGGTTGGGGATGCCCAACAATCA	2700
Db	2641	TGCTAGCTTTCTGTGTTGGTCTTAATATTGGCTAGGTTGGGGATGCCCAACAATCA	2700
QY	2701	GGTCCCTGAGATAGCTGTCTATGGGCTGATCATGTCAGAACTCTTCTCTCGGGGT	2760
Db	2701	GGTCCCTGAGATAGCTGTCTATGGGCTGATCATGTCAGAACTCTTCTCTCGGGGT	2760
QY	2761	CTGGCCCCCAAAATGCCTAACCAGGACCTTGGAAATCTACTCATCCCAAAATGATAAT	2820
Db	2761	CTGGCCCCCAAAATGCCTAACCAGGACCTTGGAAATCTACTCATCCCAAAATGATAAT	2820
QY	2821	TCCAAATGCTGTTACCAAGGTTAGGCTGTGAAGGAAGTAGAGGGTGGGGCTTCAGGT	2880
Db	2821	TCCAAATGCTGTTACCAAGGTTAGGCTGTGAAGGAAGTAGAGGGTGGGGCTTCAGGT	2880
QY	2881	CTCAACGGCTTCCCTTAACCAACCCCTCTCTCTTGGCCAGCCTGGTTCCTCCCACTTCCA	2940
Db	2881	CTCAACGGCTTCCCTTAACCAACCCCTCTCTCTTGGCCAGCCTGGTTCCTCCCACTTCCA	2940
QY	2941	CTCCCCCTACTCTCTTAGGACTGGGCTGATGAAGGCACCTGCCAAAATTTCCCTACC	3000
Db	2941	CTCCCCCTACTCTCTTAGGACTGGGCTGATGAAGGCACCTGCCAAAATTTCCCTACC	3000
QY	3001	CCCAACTTTCCCTTACCCCAACTTTCCCAAGCCTTTGTCATCTCAGCCCCCAGAGT	3060
Db	3001	CCCAACTTTCCCTTACCCCAACTTTCCCAAGCCTTTGTCATCTCAGCCCCCAGAGT	3060
QY	3061	GCAGGACAGAGAACAAGTGGGTTTTTCCCAAGCCTTTGTCATCTCAGCCCCCAGAGT	3120
Db	3061	GCAGGACAGAGAACAAGTGGGTTTTTCCCAAGCCTTTGTCATCTCAGCCCCCAGAGT	3120
QY	3121	ATATCTGTGCTTGGGGAATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGG	3180
Db	3121	ATATCTGTGCTTGGGGAATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGG	3180
QY	3181	GAGGTCTTATCTCAGGGGGGTTTTAAGTCCGTTTTGCAATAATGCTCTATTATT	3240
Db	3181	GAGGTCTTATCTCAGGGGGGTTTTAAGTCCGTTTTGCAATAATGCTCTATTATT	3240
QY	3241	TAGCGGGGTGAATATTTTATCTAAGTGAGCAATCAGAGTATAATGTTTATGTTGAC	3300
Db	3241	TAGCGGGGTGAATATTTTATCTAAGTGAGCAATCAGAGTATAATGTTTATGTTGAC	3300

QY 601 CGGCTGGCTAGCAGGGCTGTGTGCCGGATCCAGGCCCTCGAGCTGSCACTGCTCAT 660  
DB 601 CGGCTGGCTAGCAGGGCTGTGTGCCGGATCCAGGCCCTCGAGCTGSCACTGCTCAT 660  
QY 661 CCTGGGGCTGGGGCTGTGTGACCTCTGTGTGCCAGGTGTGCTTCACTCCACTGAGAGCCCT 720  
DB 661 CCTGGGGCTGGGGCTGTGTGACCTCTGTGTGCCAGGTGTGCTTCACTCCACTGAGAGCCCT 720  
QY 721 GCTCTGTGACCTCTTCCGGGACCCGAGCACCTGTGCCAGGCTACTCTGTATGCCCTT 780  
DB 721 GCTCTGTGACCTCTTCCGGGACCCGAGCACCTGTGCCAGGCTACTCTGTATGCCCTT 780  
QY 781 CATGATCAGTCTTGGGGCTGTGTGAGCTTACCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
DB 781 CATGATCAGTCTTGGGGCTGTGTGAGCTTACCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
QY 841 TGCCCTTGCCCTTACCTTGGGACCCAGGAGGAGTGCCTTGTGGCTGCTGCTGCTGCTGCTGCT 900  
DB 841 TGCCCTTGCCCTTACCTTGGGACCCAGGAGGAGTGCCTTGTGGCTGCTGCTGCTGCTGCTGCT 900  
QY 901 CTTTCTCAGCTAGCAGCCACACTGTGTGTGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
DB 901 CTTTCTCAGCTAGCAGCCACACTGTGTGTGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
QY 961 CGAGCCAGCAGAGGCTGTGGGCCCCCTTGTGTGGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
DB 961 CGAGCCAGCAGAGGCTGTGGGCCCCCTTGTGTGGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
QY 1021 CGCTTGGCTTTCGGAACCTTGGGCCCCCTTGTGTGGCCCCCTGCTGCTGCTGCTGCTGCTGCTG 1080  
DB 1021 CGCTTGGCTTTCGGAACCTTGGGCCCCCTTGTGTGGCCCCCTGCTGCTGCTGCTGCTGCTGCTG 1080  
QY 1081 CATGCCCGCACCTTCCGCGCTCTTGTGTGGCTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140  
DB 1081 CATGCCCGCACCTTCCGCGCTCTTGTGTGGCTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140  
QY 1141 GACCTTCAGCTGTGTTTACACGGATTTGTGGGCGAGGGCTGTACAGGGGCTGCCAG 1200  
DB 1141 GACCTTCAGCTGTGTTTACACGGATTTGTGGGCGAGGGCTGTACAGGGGCTGCCAG 1200  
QY 1201 AGCTGAGCGGCGCACCGAGGCGCGGAGACACTATGATGAAGCGTTCGGATGGGAGCT 1260  
DB 1201 AGCTGAGCGGCGCACCGAGGCGCGGAGACACTATGATGAAGCGTTCGGATGGGAGCT 1260  
QY 1261 GGGGCTGTGCTGAGTGCGCATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
DB 1261 GGGGCTGTGCTGAGTGCGCATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
QY 1321 GCAGCGATTGCGCACTCGAGCAGTCTATTGCGCAGTGTGGCAGTGTGGCAGTGTGGCAGTGTGG 1380  
DB 1321 GCAGCGATTGCGCACTCGAGCAGTCTATTGCGCAGTGTGGCAGTGTGGCAGTGTGGCAGTGTGG 1380  
QY 1381 CGGTGCCACATGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
DB 1381 CGGTGCCACATGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
QY 1441 GTTACCTTCTCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
DB 1441 GTTACCTTCTCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
QY 1501 GAAGCAGGTGTCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
DB 1501 GAAGCAGGTGTCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
QY 1561 CTTGATACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620  
DB 1561 CTTGATACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620  
QY 1621 GGGTCTGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680  
DB 1621 GGGTCTGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680

QY 1681 TGATGTCTCCGTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGGTGGTTCCGGGCGG 1740  
DB 1681 TGATGTCTCCGTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGGTGGTTCCGGGCGG 1740  
QY 1741 GGGCATCTCCCTGGACCTCGCCATCCTGGATAGTGCCTTCTTGTGTGCTCCAGGTGGGCGCC 1800  
DB 1741 GGGCATCTCCCTGGACCTCGCCATCCTGGATAGTGCCTTCTTGTGTGCTCCAGGTGGGCGCC 1800  
QY 1801 ATCCCTGTATTATGGCTCCATTTGTCAGCTCAGCCAGTGTGTCACTGCTATATGGTGTCT 1860  
DB 1801 ATCCCTGTATTATGGCTCCATTTGTCAGCTCAGCCAGTGTGTCACTGCTATATGGTGTCT 1860  
QY 1861 TGCCCGAGCCCTGGGTCTGGTGGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAG 1920  
DB 1861 TGCCCGAGCCCTGGGTCTGGTGGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAG 1920  
QY 1921 CGACTTGGCCAAATACTCAGCGTAGAAACTTTCAGACCATTTGGGGTGGAGGCGCTGCCT 1980  
DB 1921 CGACTTGGCCAAATACTCAGCGTAGAAACTTTCAGACCATTTGGGGTGGAGGCGCTGCCT 1980  
QY 1981 CACTGGGTCCAGAGTCCCGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040  
DB 1981 CACTGGGTCCAGAGTCCCGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040  
QY 2041 TTCTGTGCTGCCAAAGTAATGTGCT 2100  
DB 2041 TTCTGTGCTGCCAAAGTAATGTGCT 2100  
QY 2101 GCTGCACAGCTGGGGCTGGGGCTGCT 2160  
DB 2101 GCTGCACAGCTGGGGCTGGGGCTGCT 2160  
QY 2161 ACTGGAGCCCTTCAAGGGGTTTCAGTCTGGACTATACAGGAGGCGCAGAGGGCTCC 2220  
DB 2161 ACTGGAGCCCTTCAAGGGGTTTCAGTCTGGACTATACAGGAGGCGCAGAGGGCTCC 2220  
QY 2221 ATGCACCTGGAATGGGGGACTCTCAGGTGATTTACCCAGGCTCAGGGTTAACAGCTAGC 2280  
DB 2221 ATGCACCTGGAATGGGGGACTCTCAGGTGATTTACCCAGGCTCAGGGTTAACAGCTAGC 2280  
QY 2281 CTCTAGTTGAGACACACTAGAGAGGGTTTTTGGAGCTGAATAAAGCTCAGTCACTG 2340  
DB 2281 CTCTAGTTGAGACACACTAGAGAGGGTTTTTGGAGCTGAATAAAGCTCAGTCACTG 2340  
QY 2341 GTTTCCTCATCTTAAGCCCTTAACTGAGCTGCTGTTTAAATGAGCTCTTGCATGGAG 2400  
DB 2341 GTTTCCTCATCTTAAGCCCTTAACTGAGCTGCTGTTTAAATGAGCTCTTGCATGGAG 2400  
QY 2401 TTCTTAGGATGAACACTCTCTCCATGGGATTTGAACATATGACTTATTTCTAGGGAGA 2460  
DB 2401 TTCTTAGGATGAACACTCTCTCCATGGGATTTGAACATATGACTTATTTCTAGGGAGA 2460  
QY 2461 GTCTAGGGGCAACACACAGAACAGGTCCTCAGCCACAGGCTGCTCTTTTGTCT 2520  
DB 2461 GTCTAGGGGCAACACACAGAACAGGTCCTCAGCCACAGGCTGCTCTTTTGTCT 2520  
QY 2521 GATCCACCCCTTACCTTTTATCAGGATGGGCTGTTGCTCTCTGTTGCTGCTGCTGCTGCTGCT 2580  
DB 2521 GATCCACCCCTTACCTTTTATCAGGATGGGCTGTTGCTCTCTGTTGCTGCTGCTGCTGCTGCT 2580  
QY 2581 CAGACACAGGCTTAAATATTTAACTTATTTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 2640  
DB 2581 CAGACACAGGCTTAAATATTTAACTTATTTTAACTTATTTTAACTTAACTTAACTTAACTTAACT 2640  
QY 2641 TGCTAGCTTTTCTGTTGGTGTCTAATATTTGGGTAGGGTGGGGATCCCCAACATCA 2700  
DB 2641 TGCTAGCTTTTCTGTTGGTGTCTAATATTTGGGTAGGGTGGGGATCCCCAACATCA 2700  
QY 2701 GGTCCCTCAGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760  
DB 2701 GGTCCCTCAGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760  
QY 2761 CTGGCCCCCAAAATGCCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCCAAATGATAAT 2820

[illegible]

RESULT 8	
us-09-605-785-704	
; Sequence 704, Application US/09605785	
; Patent No. 6321716	
; GENERAL INFORMATION:	
; APPLICANT: Xu, Jiongchun	
; APPLICANT: Dillon, Davin C.	
; APPLICANT: Mitcham, Jennifer L.	
; APPLICANT: Harlocker, Susan L.	
; APPLICANT: Jiang, Yuqi	
; APPLICANT: Henderson, Robert A.	
; APPLICANT: Kalos, Michael D.	
; APPLICANT: Fanger, Gary R.	
; APPLICANT: Retter, Marc W.	
; APPLICANT: Stolk, John A.	
; APPLICANT: Day, Craig H.	
; APPLICANT: Vedrick, Thomas S.	
; APPLICANT: Carter, Darrick	
; APPLICANT: Li, Samuel	
; APPLICANT: Wang, Aljun	
; APPLICANT: Skeiky, Yasir A.W.	
; APPLICANT: Hepler, William	
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND	
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER	
; FILE REFERENCE: 210121.427C16	
; CURRENT APPLICATION NUMBER: US/09/605,785	
; CURRENT FILING DATE: 2000-06-27	
; NUMBER OF SEQ ID NOS: 835	
; SOFTWARE: FastSeq for Windows Version 3.0	

964	QY	964	GCCACAGAAGGGCTGTGCGGCCCTCTCTTGTGCGCCCACTGCTGTCCATGCGGGGCCG	1023
	Db	961	GCACAGAGAAGGGCTGTGCGGCCCTCTCTGTGCGCCCACTGCTGTCCATGCGGGGCCG	1020
	QY	1024	CTTGGCTTTCCGGAAACCTGGCGGCCCTGCTTTCGCCGGCTGCACAGCTGTGCTGCCGAT	1083
	Db	1021	CTTGGCTTTCCGGAAACCTGGCGGCCCTGCTTTCGCCGGCTGCACAGCTGTGCTGCCGAT	1080
	QY	1084	GCCCGCACCTGCGCCCGCTCTTTCGTGGCTGAGCTGTGAGCTGGATGSCATCATGAC	1143
	Db	1081	GCCCGCACCTGCGCCCGCTCTTTCGTGGCTGAGCTGTGAGCTGGATGSCATCATGAC	1140
	QY	1144	CTTCACGCTGTTTTACAGGATTTCTGGCGAGGGCTGTACAGGGCTGCGCCAGAGC	1203
	Db	1141	CTTCACGCTGTTTTACAGGATTTCTGGCGAGGGCTGTACAGGGCTGCGCCAGAGC	1200
	QY	1204	TGAGCGGGCACCGAGGCCGAGACACTATGATGA-----	1239
	Db	1201	TGAGCGGGCACCGAGGCCGAGACACTATGATGAAGGTAAGGCCTTGGCAGCCAG	1260
	QY	1240	-----	1239
	Db	1261	AGGCTGCTGGGAGCGCCCAACAGAGACGACACTCGGGGCTGTCTTGGGCTGGTGC	1320
	QY	1240	-----	1239
	Db	1321	TCTCCATCCTGGCCCCGACTCTCTGTTCAGGAAGTGGGATGGACCCCATCTGCATACA	1380
	QY	1240	-----	1239
	Db	1381	CGGCTTCTCATGGGTGTGGAACATCTCTGCTTCGGGTTTCAGGAAGGCCTCTGGCTGCTC	1440
	QY	1240	-----	1239
	Db	1441	TAGAGTCTGATCAGAGTCGTTGCCCAAGTTTGACAGAAGGAGGAGCTTAATCAA	1500
	QY	1240	-----	1239
	Db	1501	AGTCTAGAGGAGTGGAGGATTAAAGCTGGATTTCAGATCTGCCTGGTTCCAGCGGAG	1560
	QY	1240	-----	1239
	Db	1561	TGTGCCCTCTGCTCCCCCAACGACTTTCCAAATAATCTCACCAGCGCCTTCCAGCTCAGG	1620
	QY	1240	-----	1239
	Db	1621	CGTCTAGAAGCGTCTTGAAGCTATGGCCAGCTGCTTTGTGTGTTCCCTCTCACCCGGCT	1680
	QY	1240	-----	1239
	Db	1681	GTCTCACAGCTGAGACTCCAGGAAACCTTCAGACTACCTTCTGCTTCAGCAAGG	1740
	QY	1240	-----	1239
	Db	1741	GGCGTTGCCACATCTCTCAGGGTCAGTGGAGAACCCTAGACTCCCATTTGCTAGAGGTA	1800
	QY	1240	-----	1239
	Db	1801	GAAAGGGAGGGTGTCTGGGAGAGCGGCTGCTCCACAGAGCTCTCGTGAGCAGGTTAC	1860
	QY	1240	-----	1239
	Db	1861	CTGTGGTTCCGCTTCTCATCTCCCTGAGACTGCTCCGACCTTCCCTCCAGGCTCTGT	1920
	QY	1240	-----AGCGTTTCGGATGGCAGCGCTGGGGCTGTTCCTGTCAGT	1277
	Db	1921	CTGATGGCCCCCTCTCCCTCTGCAGCGCTTCGGATGGCAGCGCTGGGGCTGTTCCTGCAGT	1980
	QY	1278	GGSCCATCTCCCTGGTCTCTCTCTGTCATGACCGCGCTGGTGAGCGATTCGGCACTC	1337
	Db	1981	GGCCATCTCCCTGGTCTCTCTCTGTCATGACCGCGCTGGTGAGCGATTCGGCACTC	2040

QY	1338	GAGCAGCTATTATTGGCCAGTGTGCAGCTTTCCCTGTGGCTGCCGTGCCACATGCCTGT	1397
DB			
DB	2041	GAGCAGCTATTATTGGCCAGTGTGCAGCTTTCCCTGTGGCTGCCGTGCCACATGCCTGT	2100
QY			
QY	1398	CCACAGTGTGGCCGTGTGCAGCTTTCACGCCCTCACCGGTTACCTTCTCAGCCC	1457
DB			
DB	2101	CCACAGTGTGGCCGTGTGCAGCTTTCAGCGCCCTCACCGGTTACCTTCTCAGCCC	2160
QY			
QY	1458	TGCAGATCCTGCCCTACACACTTGCCTCCCTCTACCAACCGGAGAGACAGGTCTTCCCTGC	1517
DB			
DB	2161	TGCAGATCCTGCCCTACACACTTGCCTCCCTCTACCAACCGGAGAGACAGGTCTTCCCTGC	2220
QY			
QY	1518	CCAAATACCGAGGGGACACTTGGAGGTGCTAGCAGTGAGGACAGCCTGTATGACAGTTCC	1577
DB			
DB	2221	CCAAATACCGAGGGGACACTTGGAGGTGCTAGCAGTGAGGACAGCCTGTATGACAGTTCC	2280
QY			
QY	1578	TGCCAGGCCCTAAGCCTTGGAGCTCCCTTCCCTAATGGACAGTGGGTGCTGGAGGCAGTG	1637
DB			
DB	2281	TGCCAGGCCCTAAGCCTTGGAGCTCCCTTCCCTAATGGACAGTGGGTGCTGGAGGCAGTG	2340
QY			
QY	1638	GCCTGCTCCCACTTCCACCCCGCGCTTGCGGGGCTCTGCCTGTGATGTCTCCGTACGSG	1697
DB			
DB	2341	GCCTGCTCCCACTTCCACCCCGCGCTTGCGGGGCTCTGCCTGTGATGTCTCCGTACGSG	2400
QY			
QY	1698	TGGTGGTGGGTGAGCCACACGAGCCAGGTGTTCGGGCGGGGACATCGCTGGAGCC	1757
DB			
DB	2401	TGGTGGTGGGTGAGCCACACGAGCCAGGTGTTCGGGCGGGGACATCGCTGGAGCC	2460
QY			
QY	1758	TGCGCATCTGGATAGTGCTTCTGCTGTCCAGGTGGCCCCATCCCTGTTTATGGGCT	1817
DB			
DB	2461	TGCGCATCTGGATAGTGCTTCTGCTGTCCAGGTGGCCCCATCCCTGTTTATGGGCT	2520
QY			
QY	1818	CCATTGTCCAGCTCAGCAGCTGTCTACTTGCCTATATGTTGCTGCCGAGGCCCTGGGTC	1877
DB			
DB	2521	CCATTGTCCAGCTCAGCAGCTGTCTACTTGCCTATATGTTGCTGCCGAGGCCCTGGGTC	2580
QY			
QY	1878	TGGTCGCCATTTACTTTGCTACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACT	1937
DB			
DB	2581	TGGTCGCCATTTACTTTGCTACACAGTAGTATTTGACAAGAGCGACTTGGCCAAATACT	2640
QY			
QY	1938	CAGCGTAGAAAACTTCCAGCACATGGGGTGGAGGCCCTGCCTACTGGTCCCAGCTCC	1997
DB			
DB	2641	CAGCGTAGAAAACTTCCAGCACATGGGGTGGAGGCCCTGCCTACTGGTCCCAGCTCC	2700
QY			
QY	1998	CCGCTCCTGTAGCCCATGGGGCTCCCGGGCTGGCCGCAAGTGTCTGTGTCGCAAG	2057
DB			
DB	2701	CCGCTCCTGTAGCCCATGGGGCTCCCGGGCTGGCCGCAAGTGTCTGTGTCGCAAG	2760
QY			
QY	2058	TAATGTGGCTCTGTGTGCCACCTGTCTGTGAGGTGCGTAGTGCACAGCTGGGGGC	2117
DB			
DB	2761	TAATGTGGCTCTGTGTGCCACCTGTCTGTGAGGTGCGTAGTGCACAGCTGGGGGC	2820
QY			
QY	2118	TGGGGCGTCCCTCTCTCCCAAGTCTTAGGGCTGCCTGACTGGAGGCCTTTCCAAG	2177
DB			
DB	2821	TGGGGCGTCCCTCTCTCCCAAGTCTTAGGGCTGCCTGACTGGAGGCCTTTCCAAG	2880
QY			
QY	2178	GGGGTTTACGTCTGGACTTATACAGGAGGCCAGAGGCTCCATPGCACTGGAATGCGGG	2237
DB			
DB	2881	GGGGTTTACGTCTGGACTTATACAGGAGGCCAGAGGCTCCATPGCACTGGAATGCGGG	2940
QY			
QY	2238	GACTCTGCAGGTGGATTAACCCAGGCTCAGGGTTTAACAGTACGCTCCCTAGTTGAGACACA	2297
DB			
DB	2941	GACTCTGCAGGTGGATTAACCCAGGCTCAGGGTTTAACAGTACGCTCCCTAGTTGAGACACA	3000
QY			
QY	2298	CCTAGAGAAGGGTTTGGGAGCTGAATAACTCAGTCACTCGTGTTCCTCCATCTCTAAGC	2357
DB			
DB	3001	CCTAGAGAAGGGTTTGGGAGCTGAATAACTCAGTCACTCGTGTTCCTCCATCTCTAAGC	3060
QY			
QY	2358	CCCTTAACCTGCAGCTTCCTTTTAAGTAGTCTTGCATGGGAGTTTCTAGGATGAACAC	2417
DB			
DB	3061	CCCTTAACCTGCAGCTTCCTTTTAAGTAGTCTTGCATGGGAGTTTCTAGGATGAACAC	3120
QY			
QY	2418	TCCTCCATGGGATTTGAACATATG--ACTTATTTGTAGGGGAAGAGTCTCGAGGGGCAAC	2475







Db 2821 GGTGAATATTTATACCTAAGTGAGCAATCAGAGTATAATGTTATGGTGACAAAATTA 2880  
Qy 3307 AAGGCTTCTTATATGTTTAAAAA:3330  
Db 2881 AAGGCTTCTTATATGTTTAAAAA 2904

## RESULT 10

US-09-605-785-702  
; Sequence 702, Application US/09605785  
; Patent No. 6321716  
; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqi  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skelky, Yasir A.W.  
; APPLICANT: Hepler, William  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.427C16  
; CURRENT APPLICATION NUMBER: US/09/605/785  
; CURRENT FILING DATE: 2000-06-27  
; NUMBER OF SEQ ID NOS: 835  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 702  
; LENGTH: 4894  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-605-785-702

Query Match 62.8%; Score 2142.8; DB 4; Length 4894;  
Best Local Similarity 80.1%; Pred. No. 0;  
Matches 2872; Conservative 1; Mismatches 3; Indels 708; Gaps 2;

Qy 455 GGCATTGGTCCAGTCTGGGCGCTGGTCTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCAC 514  
Db 1311 GGCATTGGTCCAGTCTGGGCGCTGGTCTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCAC 1370  
Qy 515 TGGCGTGGAGCGCTATGGCGCGCCGCCCTTCATCTGGGCACTGTCTTGGGCATCTCTG 574  
Db 1371 TGGCGTGGAGCGCTATGGCGCGCCGCCCTTCATCTGGGCACTGTCTTGGGCATCTCTG 1430  
Qy 575 CTGAGCCTCTTCTCATCCCAAGGCGCGCTGGCTAGCAGGCGCTGTGCGCGGATCCC 634  
Db 1431 CTGAGCCTCTTCTCATCCCAAGGCGCGCTGGCTAGCAGGCGCTGTGCGCGGATCCC 1490  
Qy 635 AGGCGCCCTGGAGCTGGGCACTGTCTATCCCTGGGCGTGGGCTGTCTGTGGCCAG 694  
Db 1491 AGGCGCCCTGGAGCTGGGCACTGTCTATCCCTGGGCGTGGGCTGTCTGTGGCCAG 1550  
Qy 695 GTGTGCTTCACTCCACTGGAGCGCTGTCTGTGACCTCTTCCGGGACCCGGACCACTGT 754  
Db 1551 GTGTGCTTCACTCCACTGGAGCGCTGTCTGTGACCTCTTCCGGGACCCGGACCACTGT 1610  
Qy 755 CGCAGCGCTACTCTGTCTATGCTTCATGATCAGTCTTGGGGCTGCTGGGCTACCTC 814  
Db 1611 CGCAGCGCTACTCTGTCTATGCTTCATGATCAGTCTTGGGGCTGCTGGGCTACCTC 1670  
Qy 815 CTGCTGCCATTTGACTGGGACACAGTGGCCCTGGCCCTTACCTGGGACCCAGGAGGAG 874

Db 1671 CTGCTGCCATTTGACTGGGACACCAAGTGGCCCTGGCCCTTACCTGGGACCCAGGAGGAG 1730  
Qy 875 TGCTCTTTTGGCTGTCTACCCCTCATCTTCTCCTCAGCTGGGTAGCAGCCACACTGCTGGTG 934  
Db 1731 TGCTCTTTTGGCTGTCTACCCCTCATCTTCTCCTCAGCTGGGTAGCAGCCACACTGCTGGTG 1790  
Qy 935 GCTGAGGAGGACGGCTGGGCGCCACCGAGGACAGAGGGGCTGTGGGCCCCCTCTCTTG 994  
Db 1791 GCTGAGGAGGACGGCTGGGCGCCACCGAGGACAGAGGGGCTGTGGGCCCCCTCTCTTG 1850  
Qy 995 TCGCCCCACTGTGTCCATCGCGGGCGCTTGGCTTTCGGGAACCTGGGCGCCCTGCTT 1054  
Db 1851 TCGCCCCACTGTGTCCATCGCGGGCGCTTGGCTTTCGGGAACCTGGGCGCCCTGCTT 1910  
Qy 1055 CCGCGGCTGCACGAGCTGTCTGCCGATGCCCGCACCCCTGCCGCGGCTTCTCTGTTGGCT 1114  
Db 1911 CCGCGGCTGCACGAGCTGTCTGCCGATGCCCGCACCCCTGCCGCGGCTTCTCTGTTGGCT 1970  
Qy 1115 GAGCTGTGACGCTGGATGGGCACTCATGACCTTCACGCTGTTTTACAGGGATTTCTGTTGGC 1174  
Db 1971 GAGCTGTGACGCTGGATGGGCACTCATGACCTTCACGCTGTTTTACAGGGATTTCTGTTGGC 2030  
Qy 1175 GAGGGGCTGTACCAGGCGCTGCCAGAGCTGAGCGGGCACCCGAGGCGCCGAGACACTAT 1234  
Db 2031 GAGGGGCTGTACCAGGCGCTGCCAGAGCTGAGCGGGCACCCGAGGCGCCGAGACACTAT 2090  
Qy 1235 GATGA----- 1239  
Db 2091 GATGAAGGTAAGGCTTTGGCAGCCAGCAGAGGCTGTGTGGGAGCCGCCACCAGAGACG 2150  
Qy 1240 ----- 1239  
Db 2151 AACTCGGGGCTGTGTGTGGCTGGTCTCTCCATCCTGTGCCCCGACTTCTCTGTGAGG 2210  
Qy 1240 ----- 1239  
Db 2211 AAGTGGGATGAGACCCCATCTGCATACAGGCTTCTCATGGGTGTGGAACATCTCTGCT 2270  
Qy 1240 ----- 1239  
Db 2271 TGGGTTTTCAGGAAGGCTCTGGCTGTCTAGGAGTCTGATCAGAGTCTGTGCCCCAGTT 2330  
Qy 1240 ----- 1239  
Db 2331 TGACAGAGGAAAGCGGAGCTTATCAAAGTCTAGAGGAGTGGAGGAGTTAAGGCTGG 2390  
Qy 1240 ----- 1239  
Db 2391 ATTTTCAGATCTGCTGTGTCCAGCGCAGTGTGCCCTCTGCTCCCCAAGACTTTCCAA 2450  
Qy 1240 ----- 1239  
Db 2451 ATAACTCACCAGCGCTTCCAGCTCAGGCGCTCCTAGAGGCTTGAAGGCTATGGCCA 2510  
Qy 1240 ----- 1239  
Db 2511 GCTGTCTTGTGTTCCCTCTCACCCTGCTGCTCCTCAGAGTGTGAGACTCCAGGAAACCTT 2570  
Qy 1240 ----- 1239  
Db 2571 CAGACTACCTTCTCTGTGCTTTCAGCAAGGGGCGGTGGCCACATTTCTGTAGGGTCACTGG 2630  
Qy 1240 ----- 1239  
Db 2631 AAGAACCTAGACTCCCATTTGCTAGAGGTAGAAAGGGGAGGGTGTGGGAGCAGGGCTG 2690  
Qy 1240 ----- 1239  
Db 2691 GTCCACAGCAGGCTCTGTCAGCAGAGTACCTGTGTGTTCCGCTTCTCATCTCCCTGAGAC 2750  
Qy 1240 -----AGCGTTCG 1248  
Db 2751 TGCTCCGACCTTCCCTCCAGGCTCTGTCTGATGGGCCCTTCTCCCTCTGAGCGGCTCG 2810



Result No.	Score	Query			DB	ID	Description
		Match	Length	DB			
C 1	398.5	6.4	660	1	Q0B63	BHLF1 protein - hu	
C 2	373	6.0	1453	2	S21626	collagen alpha 1(I)	
C 3	364	5.8	1414	1	S23809	collagen alpha 2(I)	
C 4	358.5	5.8	1464	2	S59856	collagen alpha 1(I)	
C 5	351.5	5.6	1466	1	CGH07L	collagen alpha 1(I)	
C 6	348	5.6	1690	1	CGHUIB	collagen alpha 4(I)	
C 7	347.5	5.4	515	2	T14340	sucrose-proton tra	
C 8	344.5	5.5	1049	1	CGB07S	collagen alpha 1(I)	
C 9	343	5.5	1464	1	CGHUII	collagen alpha 1(I)	
C 10	342	5.5	1042	1	CSCHUS	collagen alpha 1(I)	
C 11	337.5	5.3	523	2	T12198	sucrose transport	
C 12	335.5	5.4	886	2	F50694	collagen alpha 1(I)	
C 13	335.5	5.4	1549	2	F48103	type VII collagen	
C 14	335.5	5.4	1843	2	S18803	collagen alpha 1(V)	

```
QY 2183 AACCCCTTGGAGGCTTCCAGTTCAGGCAGCCCTAGAGACTGGGAGAGAGAGAGGAC 2124
Db 31 GlnThrArgArgLeuProPro-----GlyThrGlyGlnArgThr 43
QY 2123 GCCCGAGCCCGAGCTGTGCAGCTACGACCTACGACAGCAGAGGCTGGCAGCAGAGACC 2064
Db 44 AlaProThrGlnValGlyLeuAlaAspAlaAlaSerProAspGluLeuGlnAspGln--- 62
QY 2063 ACATTACTTTGGCAGCAACAGAAACTGGCGCCAGCCCGCAGCCCGCTTACAG 2004
Db 63 -----AlaSerGlyAlaArgProGlyGlyGlyAsn 72
QY 2003 GAGCGGAGCTGGGAGCCAGCTGAGGAGGCGCTCCACCCCAATGTCTGAAAGTTTCT 1944
Db 73 ArgValGlyAlaGlyArg-----GlyArgProGlyThrProAla----- 85
QY 1943 ACGCTGAGTATTGGCCCAAGTCGCTTGTCAATACTACTGTGTAGCAAGTAAATGG 1884
Db 86 -----ProSerArg-----GlnSerArg--- 91
QY 1883 GCACAGACCCA-----GGCCTCGGCAG 1860
Db 92 ArgThrGlyProAlaGluGlnAlaAspHisAlaHisSerAsnProThrGlyGlyCysSer 111
QY 1859 ACACCATATAGCAGCTGACAGACTGGCTGAGCTGGACAATGGAGCCCAATAACAGGATG 1800
Db 112 AspProGlnArgSerProArgThrArgGlnAlaGlyTy-----AlaLeu 126
QY 1799 GGGCACCTGGGACAGCAGGAGGACATATCCAGATGGCAGTGCCA----- 1752
Db 127 Gly---GluGlySerAlaGlyLeu-----GlySerArgGlyProArgProHisPro 142
QY 1751 -----GGCAGATGCCCGCGCCGGAACACCGCTCGCTCGGTGG 1713
Db 143 AlaPheGlnValGlnTrpSerAlaArgAsnProGlyCysProArgThrTrpArgArgArg 162
QY 1712 GCT-----CACCCACCCACACGCTACGAGACATACAGCGCAGAGGCC 1668
Db 163 SerGlyAlaGlnArgGlyHisProPro----- 172
QY 1667 CCGAGAGCGGGTGGAGGTGGAGCAGGCCACTCCCTCCAGCACCCAGCTGCCATTAG 1608
Db 173 -----GlyAlaGly----- 175
QY 1607 GGAAGGAGCTCCAGGCTTAGGGCTGCGCAGGAGAGCTGGTCATCAGGCTGCTCACTGC 1548
Db 175 ----- 175
QY 1547 TAGCAGCTCCAGTGTCCCTCGTATTTGGGAGGAGAACACCTGCTTCCCGGTGTAGA 1498
Db 176 -----GlnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaPro 192
QY 1487 GGGAGGCCAGTGTAGGCGCAGGATCTCAGGCGTGAGAGTGAAACCGGTGAGGCGG 1428
Db 193 GlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSer----- 208
QY 1427 CTGAAGCTGTACACCGCCACACTGTGGACAGGC-----ATGTGG 1386
Db 209 ---GlyAlaThrProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAla 227
QY 1385 CACCGGCGCAGCGGAAGCTGGCAGCTGCCCAATAGACTGCTCGAGTGGCGAATC 1326
Db 228 ArgLeuProProGluArgGlnGluProArgLeuProGlnAspLeuAlaAlaGln--- 246
QY 1325 GCTGCCAGCGGCTCCATGACAGAGAGA-----CCAGGAGATGGCGCACTCAGGA 1269
Db 247 ArgCysProAlaGlyProProProProThrArgSerGlyAlaAlaAlaGlnArgHisArg 266
QY 1268 ACAGCCCGAGGCTGCCATCCGAACGCTTCATATAGTGTCTCCGGCGCTCGGTCCCG 1209
Db 267 ArgProProGlyCysPro-----ArgSerAlaArgAsnProGlyCysPro 281
QY 1208 GCTCAGCTCTGGCAGCAGCCCTGTAGAGCCCTCCGCCACGAAATCCGTGTGTAACACGG 1149
Db 282 ---ArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProPro----- 296
QY 1148 TGAAGTCTATGAGTCCATCCAGCTCAGCAGCTCAGCAGAGAGCGCGCAGGGTGC 1089
Db 297 -----ProGlyAlaGlyGln 301
QY 1088 GGGCATGCGGAGCAGCAGCTGTGTGAGCCCGGGAAGCAGGCGCCAGTTCCGGAAG 1029
Db 302 ArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrProAla 321
QY 1028 CCAAGCGGCGCGCATGACAGC-----AGTGGGCGGACA-----AGGAGG 987
Db 322 AlaProGlyProGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArg 341
QY 986 GGG-----CCGACAGCCCTCTGCTG-----GCTCGGTGGGCGCCA 951
Db 342 GlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluPro 361
QY 950 GCCTGCTCTCTCAG-----CCACGACGAGTGTGG 921
Db 362 ArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGlyProProProThr--- 380
QY 920 CTCTAGCAGGTGAGAGATGAGGTGAGCAGGCGCAAGA-----GGCAGTCTCTCT 867
Db 381 -----ArgSerGlyAlaAlaGlnArgThrHisArgArgProPro 394
QY 866 GGTGCGCCA-----GCTAGGGCGGCGGCGCAGTGTGCTCCAGTCAATGGCAGGCGAGA 813
Db 395 GlyCysProArgSerAlaArgAsnProGly-----CysProArgThrTrpArg----- 410
QY 812 GGTAGCCAGGCGCCCAAGACTGATCATGAAGCATAGACAGAGTAGGCTGCGGAC 753
Db 411 -----ArgArgSerGlyAlaGlnArg 417
QY 752 AGTGTGCGGTCCCGAAGAGGTGACAGAG-----GCAGGCGCTCCAGTG 708
Db 418 GlyHisProProProGlyAlaGlyGlnArgProSerGlyProThrGlyArgProAla 437
QY 707 GAGTGAAGCAGCTGCCACAGAAAGTCCAGCAGCCCGCCAGCAGGATGAGCAGTGCCA 648
Db 438 AlaProGlyAlaProGly-----ThrProAlaAlaProGlyProGlyGlyAlaAla 455
QY 647 GCTCCA---GGGCGCTG-----GGATCCGGCGCAGCAGCAGCCCTGCTAGC 607
Db 456 ValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProProAla 475
QY 606 CAGCGCGCTCTGGGATGAGAAAGAGGCTCAGCAGGATGCCAAGGAC----- 559
Db 476 AlaAlaArgLeuProProGluArgGlnGluProArgLeuProGlnAspLeuAlaAla 495
QY 558 -----AGTCCCGCAGATGAAGGCG 541
Db 496 GlnArgCysProAlaGlyProProProProThrArgSerGlyAlaAlaGlnArgHis 515
QY 540 CGGCGCGGCGCATAGCTCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 481
Db 516 ArgArgProPro-----GlyCysProArgSer---AlaArgAs 527
QY 480 ACCAGGCGCAGCAGCTGAGCAATGCCAGCAGCAGCAGTGTGCTGCTGCTGCTGCTGCTGCT 426
Db 527 nProGly-----CysProArgThrTrpArgArgSerGlyAlaGlnArg 542
QY 425 -----CCCCCACTTCCAGCAGCAGGCGGCGCAGATGAGTGTGCTGCTGCTGCTGCTGCT 373
Db 542 gGlyHisProProProGlyAlaGlyGln-----ArgProSerG 555
QY 372 ACCTCCAGGCGCAAGGTAGCAGGTTCAGCAGAGAGCTGGCTTTTCCGCTGCGCGCAGC 313
Db 555 yProThrGlyGlyArg-----ProAlaAlaProGly-----Al 566
QY 312 AGGCGGTCTACCCACAGCCTCTGTGACCATAGTGGCGCAGGCGGTAGGCTCAGGGGCGC 253
```

Db 566 aProGlyThrProAlaAlaProGlyPro-GlyGlyGlyAlaAlaValProSerGly----- 584  
 QY 252 GTTCAGGCACTCCAGAACTGCTGCTCGCTGCTCTCCAGAACTCGCGCTCTCTCTC 193  
 Db 585 -----AlaThrProHisProGluArg---GlySerGlyPro-----AlaAspProP 599  
 QY 192 CTTGCTGGCGCACTGCTAGGAATCAGCCAGCGCCCATTTCTGCCAGCCCTTTGGTG 133  
 Db 599 roAlaAlaAlaArgLeuProGluArg-GlnGluProArgLeuProGlnAspLeuAla 618  
 QY 132 CGGTCCTCCTCTCAGCCCTCAGCTCAGCTGCTGTGTGGGCGACCTCAGTGGGGAC 73  
 Db 619 AlaAlaGln---ArgCysProAlaGlyProProThrArgSerGlyAlaAlaAlaGln 637  
 QY 72 AGCTCTCACTCACTAGATCTTGGC-----CGAGCGCGCGGCTGTACACCGGA 25  
 Db 638 ArgThrHisArgArgProGlyCysProArgSerAlaArg---AsnProGly 654

## RESULT 2

collagen alpha 1(I) chain precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text\_change 13-Aug-1999  
 C:Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626  
 R:Li, S.W.; Khillan, J.; Prockop, D.J.  
 Matrix Biol. 14, 593-595, 1994  
 A:Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I  
 A:Reference number: S57243  
 A:Accession: S57243  
 A:Molecule type: mRNA  
 A:Residues: 1-1453 <LIS>  
 A:Cross-references: EMBL:U08020; NID:9470673; PIDN:AA88912.1; PID:9470674  
 R:Meisner, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
 Biochim. Biophys. Acta 1089, 241-243, 1991  
 A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.  
 A:Reference number: S16176; MUID:91274355; PMID:2054384  
 A:Accession: S16374  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1442-1453 <MET>  
 A:Cross-references: EMBL:X57981; NID:950484; PIDN:CAA41046.1; PID:950485  
 R:French, B.T.; Lee, W.H.; Maul, G.G.  
 Gene 39, 311-312, 1985  
 A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.  
 A:Reference number: A23982; MUID:86137403; PMID:3841523  
 A:Accession: A23982  
 A:Molecule type: mRNA  
 A:Residues: 518-1128 <FRE>  
 A:Cross-references: GB:M14423; NID:g192261; PIDN:AAA37333.1; PID:g192262  
 R:Monson, J.M.; Friedman, J.; McCarthy, B.J.  
 Mol. Cell. Biol. 2, 1362-1371, 1982  
 A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for  
 A:Reference number: I49559; MUID:83141374; PMID:6298597  
 A:Accession: I49559  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 735-1130 <RES>  
 A:Cross-references: GB:M17491; NID:g192263; PIDN:AAA37334.1; PID:g192264  
 R:Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984  
 A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads  
 A:Reference number: I49557; MUID:84170331; PMID:6324198  
 A:Accession: I49557  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-25 <RE2>  
 A:Cross-references: GB:K01688; NID:g192246; PIDN:AAA37330.1; PID:g553881  
 R:Fenton, S.P.; Lemande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.  
 Biochim. Biophys. Acta 1216, 469-474, 1993  
 A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.  
 A:Reference number: S39789; MUID:94092741; PMID:8268229  
 A:Accession: S39789

A:Molecule type: DNA  
 A:Residues: 1-80, 'E', '82-105, 'D', 107-185; 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 122  
 R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.  
 Mol. Cell. Biol. 14, 5950-5960, 1994  
 A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indi  
 A:Reference number: I48300; MUID:94344105; PMID:8065328  
 A:Accession: I48300  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-80, 'E', '82-105, 'D', 107-147 <REF>  
 A:Cross-references: EMBL:X54876; NID:950486; PIDN:CAA38657.1; PID:950487  
 C:Genetics:  
 A:Gene: COL1A1  
 A:Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 932/3; 968/3; 1004/3; 1022/3; 10  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple hel  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-151/Domain: amino-terminal propeptide #status predicted <PRO>  
 F:30-89/Domain: von Willebrand factor type C repeat homology <VWC>  
 F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>  
 F:1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

## Alignment Scores:

Pred. No.: 5,45e-15 Length: 1453  
 Score: 373.00 Matches: 285  
 Percent Similarity: 31.98% Conservative: 47  
 Best Local Similarity: 27.46% Mismatches: 371  
 Query Match: 5.99% Indels: 336  
 DB: 2 Gaps: 55

US-09-759-143-110 (1-3410) x S21626 (1-1453)

QY 2594 TGCTGTGT-----CTCTGTGTATGTC-----AACAGAGGACCAAC 2559  
 Db 52 CysileCysHisasnGlyThrAlaValCysAspValGlnCysasnGluGluLeuAsp 71  
 QY 2558 AGCCACACTCTGATAAAGTAAGAGGGGTGGATCAGCAAAAGACAGTCTGT--- 2502  
 Db 72 CysProasnProGlnArgArg---GluGlyGly-----CysCysAla 84  
 QY 2501 -----GGGCTAGGGGACCTGTGT 2484  
 Db 85 PheCysProGluGluTyrValSerProasnSerGluAspValGlyValGluGlyPro--- 103  
 QY 2483 TCTGTGTGTCTGCCCTCAGGACTTCCCTACAAATAGTCATATGTTCAATCCCAT 2424  
 Db 104 -----LysGlyGly 106  
 QY 2423 GGAGGAGTCTTTCATCTAGAACTCCCATGCAAGAGCTACATTAACGAAGCTGCAGT 2364  
 Db 107 ProGlyProGlnGlyProArgGly-ProValGlyPro----- 118  
 QY 2363 TAGGGGCTTAGAGATGGGAACACAGGTGACTGAGTTTATTTCAGTCTCCCAAAACCCCT--- 2306  
 Db 119 -ProGly---ArgAspGlyIleProGlyGlnProGlyLeuProGlyProProGlyPro 137  
 QY 2305 ----TCTCTAGCTGTCTCAACTAGGAGGCTAGCTGTGTAACT----- 2267  
 Db 137 OGlyProProGlyProProGlyLeuGlyGlyAsnPheAlaSerGlnMetSerTyrGlyTy 157  
 QY 2266 -----CTGAGCTGGTGAATCACTGCAGAGTCCCGCATTCACAG 2226  
 Db 157 rAspGluLysSerAlaGlyValSerValProGlyProMetGlyProSer----- 173  
 QY 2225 TCATGGAGCCCTTCTGGCTCTTAACTTCCAGACTGAAACCCCTTGGAGAGGCT 2166  
 Db 174 -----GlyProArgGlyLeuProGlyProProGlyAlaProGlyPro---GlnGlyPh 190  
 QY 2165 CCAGTCAGCAGCCCTCAGACTGGGGAGAG-----AGGAG 2130  
 Db 190 eGln---GlyProGlyGluProGlyGlySerGlyProMetGlyProAr 209  
 QY 2129 AGGGACGCCGCCAGCCCCCAGCTGTGTACGCTACGCACTCAGCAGCAGCAGGTCGACAG 2070



QY	252	GTTCAGGCAC	TCCAGAACTGCTTCGTCTCGGCTCTCTCCAGAACTGCGGCTCTCCTC	193
Db	844	snValGlyAla	ProGly-----ProLysGlyProArgGlyAlaAlaGlyPro	860
QY	192	CTTGTCGCGGCAACTGCCTAGGAATCAGCAGCGGCCCATTTCTGCCAGCCCTTGGTG	133	
Db	860	ro-GlyAlaThr	-----GlyPheProGlyAlaAlaGlyArgValGlyProGly	876
QY	132	CGGTCCAGCTTCTCAGCCCATGCTCAACACTGCTGCTGTGGGCACCTCAGT	79	
Db	877	ProSerGlyAsnAlaGlyProProGlyProProGlyProValGlyLysGluGlyLys	896	
QY	78	-----GGGACAGCTCTCATCTCAGATCCTGGCCGA	46	
Db	897	GlyProArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProProGlyPro	916	
QY	45	-----GGCGCGGGCTGTCAACCGGAGCC	22	
Db	917	ProGlyProAlaGlyGluLysGlySerProGlyAla	928	

### RESULT 3

collagen alpha 2(I) chain homolog - sea urchin (Strongylocentrotus purpuratus)  
C:Species: Strongylocentrotus purpuratus (purple urchin)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S23809  
R:Exposito, J.Y.; d'Alessio, M.; Sollursh, M.; Ramirez, F.  
J. Biol. Chem. 267, 13559-13562, 1992  
A:Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) collagen  
A:Reference number: S23809; MUID:92348411; PMID:1639795  
A:Accession: S23809  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1414 <EXP>  
A:Cross-references: EMBL:M92040; NID:g161435; PIDN:AAA30035.1; PID:g161436  
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology  
C:Keywords: collod coll; extracellular matrix; glycoprotein; trimer; triple helix  
F:1207-1414Domain: fibrillar collagen carboxyl-terminal homology <FC>







Db 464 lyLys-AspGlySerProGlyGluPro--GlyAlaAsnGlyLeuProGlyAlaAlaGly 483 : : : : :  
QY 1995 AGCTGGGACCCAGT-----GAGGCAG 1975 : : : : :  
Db 483 luArgGlyProSerGlyPheArgGlyProAlaGlyProAsnGlyIleProGlyGluLys 503 : : : : :  
QY 1974 GCCTCCCA-----CCCCAATGTCTGGAGTTTCTACGCTGAGTATTG 1930 : : : : :  
Db 503 lyProProGlyGluArgGlyGlyProGlyProAlaGlyPro-----ArgGlyValAlaG 521 : : : : :  
QY 1929 GCCAAGTCGCTCTGTCAAAATACCTACCTGTGTAGCAAAAGTAATAGGCGACACACCCAGG 1870 : : : : :  
Db 521 lyGluProGly-ArgAspGlyThr-Pro-----GlyGlyProGly 533 : : : : :  
QY 1869 CTGCGGCGACACCATATAGCAGTACAGACTGGCTGAGCTGACAAATGGAGCCCAT 1810 : : : : :  
Db 533 lyLeuArgGlyMetProGlySerProGlyGlyProGlyAsnAspGlyLysProGlyProPr 553 : : : : :  
QY 1809 AACAGGGATG-----GGCCACCTGGGACACAGGAGGCAC----- 1773 : : : : :  
Db 553 oclySerGlnGlyGluSerGlyArgProGlyProProGlyProSerGlyProArgGlyG 573 : : : : :  
QY 1772 -----TATCAGGATGCGAGGTCCAGGACAGATCCCGCCCGGAC- 1730 : : : : :  
Db 573 nProGlyValMetGlyPheProGlyProLysGly-AsnAspGlyAlaProGlyLysAsnG 593 : : : : :  
QY 1729 -----CACCTCGCTCGTGGCTCACCCACCACACACGACGACGAGCATCA 1681 : : : : :  
Db 593 lyGluArgGlyGlyProGlyGlyProGlyLeuPro----- 604 : : : : :  
QY 1680 CAGGACAGAGCCCGCCAGAGCGGGTGGAGTGGGAGCAGGCCACTCCCTCCACACCC 1621 : : : : :  
Db 605 -----GlyProAlaGlyLys-AsnGlyGluThrGlyProGlnGlyProProGlyPro 621 : : : : :  
QY 1620 AGCTGTCATAGG-----AAGGGA 1600 : : : : :  
Db 622 ThrGlyProAlaGlyAspLysGlyAspSerGlyProProGlyProGlnGlyLeuGlnGly 641 : : : : :  
QY 1599 GCTCAGGCTTAGG----- 1585 : : : : :  
Db 642 IleProGlyThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLys 661 : : : : :  
QY 1584 -----CCTGCGAGAGAGTGGTGCATCAGCTGTCTCTCACTG 1549 : : : : :  
Db 662 GlyGluValGlyAlaProGlyAlaProGly-GlyLysGlyAspSerGlyAlaProGlyG 681 : : : : :  
QY 1548 CTAGCACCTCCAGTGTCCCTCGGTATT-----TGGCGAGGAACACCTGC 1504 : : : : :  
Db 681 uArgGlyProProGlyThrAlaGlyIleProGlyAlaArgGlyGlyAlaGlyProProG 701 : : : : :  
QY 1503 TTCTCCCGGTGTAGAGG-----AGGCCAGTGTGTAGGGCAGGATCTGC 1459 : : : : :  
Db 701 yProGluGlyGlyLysGlyProAlaGlyProProGlyProProGlyAlaSerGlySerPr 721 : : : : :  
QY 1458 AGGCGTGAAGGTGAACCGGTGAGCGGCGGTGAAGTGTCCACCAGCCACACTGTGG 1399 : : : : :  
Db 721 oGlyLeu-GlnGlyMetProGlyGluArgGly----- 731 : : : : :  
QY 1398 GACAGGCATGTGCGCACCGGCGACGACAGGGAAGTGCACACTGGCCAAATAGACTGCT 1339 : : : : :  
Db 732 -----GlyProGlySerProGlyProLysGlyGluLysGlyGluProGlyGlyA 748 : : : : :  
QY 1338 CGAGTGCCGAATCGCTGCA-----CCAGCCGCT----- 1311 : : : : :  
Db 748 laGlyAlaAspGlyValProGlyLysAspGlyProArgGlyProAlaGlyProIleGlyP 768 : : : : :  
QY 1310 --CCATGACAGAGAGAACACAGGAGATGGCGCACTCCAGGACAGCCCGGCTGC- 1254 : : : : :  
Db 768 roProGlyProAlaGlyGlnProGlyAsp---LysGlyGluGlyGlySerProGlyLeuP 787 : : : : :  
QY 1253 -----CCATCCGACGCTTTCATCATAGTGTCTCCGGGC 1220 : : : : :  
787 roGlyIleAlaGlyProArgGlyGlyProGlyGluArgGlyGluHis-----GlyP 804 : : : : :  
1219 CTGGTCCCGGCTCAGCTCTGGGCACGCCCTGTGTACAGCCCTCGCCACAAATCCGT 1160 : : : : :  
804 roProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluPro----- 820 : : : : :  
1159 GTAACACAGCGTGAAGTCTAGTCCCATCCAGCTGCACAGCTCAGCCACCAAGAGCCG 1100 : : : : :  
820 ----- 820 : : : : :  
1099 GCCCAGGTCGGGGCATCGGCAGCACAGCTGGTCAGCCGGGGAAGCAGGCGGCCCA- 1041 : : : : :  
821 -----GlyAlaLysGlyGluArgGlyAlaProG 830 : : : : :  
1040 -----GTTCCGAAAGCCAAAGCGGCCCGGATGACAGCAGTGGGGCGACAAGA 989 : : : : :  
830 lyGluLysGlyGluGlyProProGlyProAla----- 841 : : : : :  
988 GGGGGCGCAGCCCTCTCTGCTGGTGGGGCCAGCGTCCCTCCTCAGCCACCAG 929 : : : : :  
842 --GlyProThr-----GlySerSerGlyProAlaGlyProGlyProGlyProGlnG 957 : : : : :  
928 CAGTGTGGCTGCTACGCGGTGAGGAAGATGAGGGTGAGCAGGCCCAAGAGGACACT- 873 : : : : :  
857 lyVal-----LysGlyGluArgGlySerProG 866 : : : : :  
872 --CCTCCTGGTCCCGAGGTAGGGGCCAGGCGACTGGTGTCCCATCTAATGCCAGGCAG 815 : : : : :  
866 lyProGlyThrAlaGlyPheProGlyGlyArgGlyLeuProGlyProProGlyAsnA 886 : : : : :  
814 GAGGTAGCCCGCAGCAGCCCAAGACTGATCATGAAGGCATACAGAGTAGGCTGGCG 755 : : : : :  
886 snGlyAsnProGlyProProGlyPro-----SerGlyAlaProGlyL 900 : : : : :  
754 ACAGTGTGTCG-----GTCGCCGAAGAGGTCTCAGAGAGCAG 719 : : : : :  
900 ysAspGlyProProGlyProAlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaG 920 : : : : :  
718 GGCCTCAGTGGAGTGAAGCACACCTGGGCCACAGAGTCCAGAGCCCGCCAGGAT 659 : : : : :  
920 lyProLysGlyAlaGlyGlnProGlyGlyLysGlyPro-----ProGly- 935 : : : : :  
658 GAGCAGTGCCAGCTCCAGGGCGCTGGATCCG---GGCACAGCAGCCCTGCTAGCCAGCC 602 : : : : :  
936 --AlaGlnGlyProProGlySerProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaA 955 : : : : :  
601 GGCCCTTGGATGAGAAGAGGTCTCAGCAGATGCCCAAGGACAGTCCCATGATGAAGG 542 : : : : :  
955 rgGlyLeuAlaGlyProProGlyMetProGly---ProArgGlySerProGlyProGlnG 974 : : : : :  
541 CCGCGCGGCGCATAGC-----GTCCACGCGCTGCTCAGCTGAGCTAGGAGCGG 488 : : : : :  
974 lyIleLysGlyGluSerGlyLysProGlyAlaSerGlyHis-----AsnG 989 : : : : :  
487 GACACAGA-----CCAGGCGCCAGCAGTGGACCAATGCCAGCACCATTGTCATGAAC 434 : : : : :  
989 lyGluArgGlyProProGlyPro-----G 997 : : : : :  
433 CTCCTCTACCCCACTTCCAGCAGCAGAGGGCGGCACATAGTGTGCTGCTGGGCCCAACA 374 : : : : :  
997 lnGlyLeuProGlyGlnProGlyThrAlaGly-----\*GluProGlyA 1011 : : : : :  
373 CACCTCCAGGCCAAAGGTTACAGGTTGACCAAGAGCTGGGCTTTCCGGTGCC----- 318 : : : : :  
1011 rgAspGlyAsnProGlySerAspGlyGlnProGlyArgAspGlySerProGlyGlyLysG 1031 : : : : :  
317 -----GCAGAGCGGCTCACCCACAGCCTCTGGACCATGAGTGG--- 278 : : : : :  
1031 lyAspArgGlyGluAsnGlySerProGlyAlaProGlyAlaProGly-HisProGlyPro 1050 : : : : :  
277 CCAGCGGGTAGGCTCAGGGGCGGTTCAGGCACTCCAGCAACTGCTCTCGGCTCT 218 : : : : :  
1051 ProGlyProValGlyProSerGlyLysSerGly-----AspArgGlyGluThrGlyPro 1068 : : : : :  
Db









Db 308 yProArglyAspProGlySerTyrgly- ||||| ||||| SerProGlyPheProGlyLe 324  
QY 2711 CTCAGGGACCTG---ATTGTGGGATCC- ||||| ||||| 2584  
Db 324 uLysGlyLeuGlyLeuValGlyAspProGlyLeuPheGlyLeuLeGlyProLysG1 344  
QY 2683 -----CCACCC- ||||| ||||| TACCATAATATTAG 2664  
Db 344 yAspProGlyAsnArgGlyHisProGlyProGlyValLeuValThrProLeuPr 364  
QY 2663 ACACCAACACAGAAAGCTAGCAATGATTCCTCTACT- ||||| ||||| 2624  
Db 364 oLeuLysGlyProProGlyAspProGlyPheProGlyArgTyrglyGluThrGlyAspVa 384  
QY 2623 -----TTGTTAAATAAATAAGTTAAATATTAAATCCCTGTGT 2586  
Db 384 lGlyProProGlyProProGlyLeuLeuGlyArgProGlyGlu- ||||| ||||| AlaCysAl 401  
QY 2585 CTCTGTGATGCCAAGACAGACAGGCCAC- ||||| ||||| ATCCTGATAAAAGGTAA 2535  
Db 401 aglyMetileGlyProProGlyProGlnGlyPheProGlyLeuProGlyLeuProGlyG1 421  
QY 2534 GAGGGGGTGGATCAGCAAAAGACAGCTGCTGGCTGAGGGACCTGTTCTTGTGTG 2475  
Db 421 uAlaGlyLeuProGlyArgProAspSerAlaProGly- ||||| ||||| 433  
QY 2474 TTGCCCTCAGGACTCTTCCCTTACAATAAGTATATGTTCAATCCCATGGAGAGTG 2415  
Db 433 ----- 433  
QY 2414 TTTTCATCTAGAACTCCCATGTCAGAGCTACATTAAACGAAGCTCAGGTTAAGGGCT 2355  
Db 433 ----- 433  
QY 2354 TAGAGATGGAAACAGGTGACTGAGTTATTTCAGCTCCCAAAACCTTCTCTA---GG 2298  
Db 434 LysProGlyLysProGly- ||||| ||||| SerProGlyLeuProGlyAlaProG1 448  
QY 2297 TGTGCTCAACTA---GGAGGTAGCTGTTAAACCTGAGCCTGGGTAATCCACCTGCAGA 2241  
Db 448 yLeuGlnGlyLeuProGlySerSerValileTyrcysSerValGlyAsnProGlyProG1 468  
QY 2240 GTCCCGCATTCAGTCATGAGCCCTCTGCTGCTCCCTGATATAGTCCAGACTCAAAC 2181  
Db 468 nGlyLeuLysGlyLysValGlyProProGlyArgGly- ||||| ||||| ProLysGlyGlu 486  
QY 2180 CCCCTTGGAGGCTCCAGTCAGGCCTCAGACTGGGAGAGAGAGAGAGAGGCC 2121  
Db 486 sGlyAsnGluGlyLeuLysAlaCysGluProGlyPro- ||||| ||||| MetGlyProPr 502  
QY 2120 CCAGCCCCAGCTGTGACCTACGACCTCAGCAGCAGGCTGGCAGCAGAGAGCCACA 2061  
Db 502 oGlyProProGlyLeuProGlyArg- ||||| ||||| 2018  
QY 2060 TTACTTTGCAACACAGAACTGGGCCA---GCCCGGAGGCC- ||||| ||||| 2018  
Db 521 yTrpLeuGlyThrLysGlyAspProGlyProGlyAlaGluGlyProProGlyLeuPr 541  
QY 2017 -----CATGGGGCTAAC- ||||| ||||| 2006  
Db 541 oGlyLysHisGlyAlaSerGlyProProGlyAsnLysGlyAlaLysGlyAspMetValva 561  
QY 2005 -----AGAGGGGAGCTGGGCCAGTCCAGTCCAGGCCCTTCCA- ||||| ||||| 1967  
Db 561 lserArgValLysGlyHisLysGlyGluArgGlyProAsp- ||||| ||||| GlyProGlyPh 579  
QY 1966 -CCCCAATGTGTGCAAGTTTCTACGCTGAGTATTTGGCCCAAGTCCGCTCTTGTCAATA 1908  
Db 579 eProGlyGlnProGlySerHisGlyArg- ||||| ||||| AspGlyHisAlaGly- ||||| ||||| 593  
QY 1907 CTACCTGTGTAGCAAAAGTAAATGGCGACCA- ||||| ||||| GA 1876

594 -----GluLysGlyAspProGlyProProGlyAspHisGluAspAlaTh 608  
1875 CCAGCCCTCCGACACACACATATAGCAGTGCACAGCTGTGAGTGGACAAATGGAG 1816  
608 rProGlyGlyLysGlyPheProGlyProLeuGlyProProGlyLysAlaGlyProValG1 628  
1815 CCATATAACAGGATGGGCCACCTGGGACAGCAGGAAGCAC- ||||| ||||| 1773  
628 yPro-----ProGlyLeuGlyPheProGlyProProGlyGluArgGlyHisProGlyValPr 647  
1772 ----TATCCAGGATGGCAGAGTCCAGGACAGATGCCCGCCCGGAA- ||||| ||||| 1731  
647 oGlyHisProGlyValArgGlyProAspGlyLeuLysGlyGlnLysGlyAspThrIleSe 667  
1730 -----CCACCTGGCC 1720  
667 rCysAsnValThrTyProGlyArgHisGlyProProGlyPheAspGlyProProGlyPr 687  
1719 TCGGTG- ||||| ||||| GG 1712  
687 oLysGlyPheProGlyProGlnGlyAlaProGlyLeuSerGlySerAspGlyHisLysG1 707  
1711 CTCACCCACACACACGTACGAGACATCACAGCAGAGGCCCGCCGAGAGCGCGGT- 1654  
707 yArgProGlyThrProGlyThr- ||||| ||||| AlagluLeuProGlyProProGlyPh 723  
1653 -----GGAGGT- ||||| ||||| GGAGCAGGCCACTG-CC 1631  
723 eArgGlyAspMetGlyAspProGlyPheGlyGlyLysGlySerSerProValG1Pr 743  
1630 TCCACACCCAGCTGCTCCATTAGG- ||||| ||||| AAGGAGCTCCAGC- 1591  
743 oProGlyProProGlySerProGlyValAsnGlyLysGlyLeuProGlyAspProAl 763  
1590 -----TTAGGG- ||||| ||||| CCTGGCAGAGCT 1571  
763 aPheGlyHisLeuGlyProProGlyLysArgGlyLeuSerGlyValProGlyLysLys-G 783  
1570 GGTATCAGGCTGCTCCTCAGTCCAGTCCAGTGTCCCTCGGTATT- 1521  
783 lyProArgGlyAspProGlyCysProGlyAlaGluGlyProAlaGlyLeuProGlyPheL 803  
1521 ----- 1521  
803 euGlyLeuLysGlyProLysGlyArgGluGlyHisAlaGlyPheProGlyValProGlyP 823  
1521 ----- 1521  
823 roProGlyHisSerCysGluArgGlyAlaProGlyLysProGlyGlnProGlyLeuProG 843  
1520 --TGGCAGGAGAACACCTGCTCTCCCGTGTAGAGGAGGCCA- ||||| ||||| GTGTGTAGG 1469  
843 lyTyProGlySerProGlyAlaProGlyLysGlyGlnProGlyAspValGlyProp 863  
1468 CAGGATCTGAGGCTGAGAAGGTGAACCCGCTGAGGCGGCTGCAAGCTCTCACCAGGC 1409  
863 roGlyProAlaGlyMetLys- ||||| ||||| GlyLeuProGlyLeuProGlyArgP 878  
1408 CACACTGGGACAGGATGTGGACCGGACGACAGGAGGAGGAGGAGGAGGAGGAGGAG 1349  
878 roGlyAla- ||||| ||||| HisGlyProPro- ||||| ||||| GlyLeuProGlyLysProG 891  
1348 ATAGACTGCTGAGTCCGATCGCTGACCCAGCCGCTCCA- ||||| ||||| TGACAGAGAGAGAG 1292  
891 lyProPheGlyAspAspGlyLeuProGlyProProGlyProGlyProLysGlyProArgGlyLeuP 911  
1291 CAGGGAGATGGCAGCTGAGGAGAACAGCCCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1232  
911 roGly- ||||| ||||| PheProGlyPheProGlyGluArgGlyLysProG 924  
1231 GTCTCTCCGGGCTCGGTCCCGGCTCAGCTGTGGGACGCGCTGTACAGCCCTCGGCC 1172  
924 lyAlaGluGly- ||||| ||||| CysProGlyAlaLys- ||||| ||||| GlyLeuProGlyLysGlyMetLys 941



[illegible]

QY 452 -----GCACCATGTCAT 440  
Db 1274 laProGlyProProGlyLeuProGlySerValAspLeuLeuArgGlyGluProGlyAspC 1294  
QY 439 GAACTTCTCTACCCCGCATTCACAGCAGCAGAGCGGCACATAGGTGATGCTCGTCGCGC 380  
Db 1294 ysGlyLeuProGlyProProGlyProProGly-----p 1305  
QY 379 CAACACACACTCCAGGCCAAAG-----TTAGCAGGTTGACCAAG 341  
Db 1305 roProGlyProProGlyTyrLysGlyPheProGlyCysAspGlyLysAspGly-GlnLys 1324  
QY 340 CAAGAGCTGGCTTCGGTGGCGGAG---CAGCGCGCTCACCCACAGCTCTGGGACCAT 284  
Db 1325 GlyProMetGlyPheProGlyProGlnGlyProHisGlyPheProGlyProGlyGlu 1344  
QY 283 AGTGG---CCAGGCGGTAGGCTCAGGGGGCGTTTCAGGCACTCCAGAACTGCTTCGT 227  
Db 1345 LysGlyLeuProGlyProProGlyArgLysGlyProThrGlyLeuPro----- 1360  
QY 226 CTCGGCTCTCTCCAGAAAGCTGCGGCGCTCTCTCTCTTC-----TCCCGCCAA 179  
Db 1361 -----GlyProArgGlyGluProGlyProProAlaAspValAspAspCysProArg 1377  
QY 178 CTCGCTTAGGAATCAGCAGCGCGCCCTTCTGCCAGCCCTTTGGTCCGCTCCAGCTTCT 119  
Db 1378 IlePro-GlyLeuProGlyAlaProGlyMetArgGlyPro----- 1390  
QY 118 CAGCCCATGCTCAACACACTGCTGCTGGGGCACCTCAGTGGGGACACGCTCTCATCACTC 59  
Db 1391 -----GluGlyAlaMetGly---LeuProGly-----MetArgG 1401  
QY 58 AGATCTTGGCGAGCGCGGTGTTCACCGGACGCGCGTGCAGG 10  
Db 1401 yProGlyProGlyCysLysGlyGluProGlyLeuAspGlyArgArg 1417  
RESULT 7  
T14340  
sucrose-proton transport protein - carrot  
N:Alternate names: sucrose/H+ symporter protein  
C:Species: Daucus carota (carrot)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: T14340  
R:Shakya, R.; Sturm, A.  
Plant Physiol. 118, 1473-1480, 1998  
A:Title: Characterization of source- and sink-specific sucrose/H+ symporters fr  
A:Cross-references: EMBL:Y16768; NID:Q2969883; PIDN:CAA76369.1; PID:q2969884  
A:Reference number: Z17991; MUID:99063785; PMID:9847123  
A:Accession: T14340  
A>Status: preliminary; translated from GB/EMBL/DBDJB  
A:Molecule type: mRNA  
A:Residues: 1-515 <SHA>  
A:Cross-references: EMBL:Y16768; NID:Q2969883; PIDN:CAA76369.1; PID:q2969884  
A:Experimental source: cultivar Nantaise; root

A; Note: SUT2  
C: Superfamily: common tobacco sucrose transport protein

Alignment Scores:	
Pred. No.:	1.9e-13
Score:	347.50
Percent Similarity:	41.78%
Best Local Similarity:	25.94%
Query Match:	5.41%
DB:	2
Length:	515
Matches:	131
Conservative:	80
Mismatches:	215
Indels:	79
Gaps:	17

US-09-759-143-110 (1-3410) x T14340 (1-515)

QY 332 CAGCTTCTGCTGGTCAACCTGCTAAACCTTTGGCCCTGGAGTGTGTTGGCCCGACGCATC 391  
 Db 34 LysLeuValLeuValAlaAlaAlaAlaAlaAlaAlaGlyValGlnPheClyTrpAlaLeuGlnLeu 53  
 QY 392 ACCTATGTGCCGCCCTCTGCTGCTGGAACTGGGGGTAGAGGAGCAAGTTTCATGACCATGGTG 451



Db 54 SerLeuLeuThrProTyrValGlnLeuLeuGlyIleProHisLysTrpAlaAlaTyrIle 73  
 QY 452 CTGGCATTTGGTCCAGTGTGGCCCTGTGTCTGCCCTCTAGGCTACCCAGTGC 511  
 Db 74 TrpLeuGlyGlyProIleSerGlyMetLeuValGlnProIleValGlyTyrTyrSerAsp 93  
 QY 512 CACTGGCGTGGAGCGTATATGGCCCGCCCGCTTCTCATCTGGGCACTGTCTTGGGCATC 571  
 Db 94 HisCysGlnSerSerPheGlyArgArgProPheIleAlaSerGlyAlaGlyCysVal 113  
 QY 572 CTGCTCAGGCTCTTCTCTATCCCAAGGCGCGCTGGCTAGCAGGCGTG 619  
 Db 114 AlaIleSerValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 130  
 QY 620 -----CTGTGCGCGGATCCAGGCCCTGGAGCTGGCAGCTCTCTCTCTCTCTCTCTCT 670  
 Db 131 GlyAspAspMetSerLysThrLeuLysProArgAlaValThrValPheValIleGlyPhe 150  
 QY 671 GGGCTGTGGACTTCTGTGGCAGGTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 730  
 Db 151 TrpIleLeuAspValAlaAsnMetLeuGlnGlyProCysArgAlaLeuLeuAlaAsp 170  
 QY 731 CTCTTCCGG---GACCCGACCATCTGTCCGCGGCTACTCTGTCTATGCTCTCATGATC 787  
 Db 171 LeuCysSerGlyAspThrArgMetArgSerAlaAsnAlaPheTyrSerPheMet 190  
 QY 788 ACTCTTGGGGCTGCTGGGCTAC-----CTCCTG 817  
 Db 191 AlaValGlyAsnIleLeuGlyTyrAlaAlaGlySerTyrAsnAsnLeuTyrLysLeuPhe 210  
 QY 818 CTGGCATTTGGTCCAGTGTGGCCCTGTGTCTGCCCTCTAGGCTACCCAGTGC 877  
 Db 211 ProPheSer-----LysThrHisAlaCysAspLeuTyrCysAlaAsnLeuLysSerCys 228  
 QY 878 CTCTTTGGCTGCTCACCT 937  
 Db 229 PheIleIleSerIleAlaLeuLeuIleIleIleIleIleIleIleIleIleIleIleIleIle 948  
 QY 938 GAGGAGGAGCGCTGGGCGCCCGCCAGCAGCAGGAGGCTGTGCGGCCCTCTCTCTCTCTCT 997  
 Db 249 GluAsnSer-----GlyProProAspAlaAspAlaAlaGluGluPro-----263  
 QY 998 CCCCCT 1057  
 Db 264 -----ProSerSerGlyLysIleProVal-----PheGlyGluLeuLeuGly 277  
 QY 1058 CGGTGTCCACCATCTGT 1117  
 Db 278 AlaLeuLysAspLeu-----ProArgProMetLeuLeuLeuLeuLeuLeuLeuLeuLeu 293  
 QY 1118 CTGTGTCTGT 1177  
 Db 294 CysLeuAsnTrpIleAlaTrpPheProPheIleLeuPheAspThrAspTrpMetGlyArg 313  
 QY 1178 GGGCTGTACACAGGCGGTGCCAGAGCTGTGAGCGGCGCCAGGCGCCGAGCAGCATGAT 1237  
 Db 314 GluIleTyrGlyGlyThr-----AlaGlyGlnGly-----LysLeuTyrAsp 327  
 QY 1238 GAAGCGGT 1297  
 Db 328 GlnGlyValArgAlaGlyAlaLeuGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 347  
 QY 1298 TCTGT 1357  
 Db 348 SerIleAlaValGluTyrLeuValArgGlyValGlyValLysIle-----LeuTrpGly 366  
 QY 1358 GTGCGAGCTTTCTCTGT 1396  
 Db 367 PheValAsnPheIleLeuAlaIleGlyLeuValMetThrValValSerLysValAla 386  
 QY 1397 -----TCCACAGTGTG-----GCCGTGGTGACA 1420  
 Db 387 GlnHisGlnArgGluHisSerAlaAsnGlyGlnLeuLeuProProSerAlaGlyValLys 406

QY 1421 GCTTCAGC-----GCCCTCACCAGGTTACCTTCTCAGCCCTGCAGATCCTG 1468  
 Db 407 AlaGlyAlaLeuSerLeuPheSerIleLeuGlyIleProLeuSerIleThrTyrSerIle 426  
 QY 1469 CCCTACACACTGGCTCCCTCTACACCGGAGAGAGAGGTGTCTCTCCCTCCCAATACCGA 1528  
 Db 427 ProPheAlaLeuAlaSerIleTyrSerSerGlySerGlyAlaGlyGlnGlyLeuSerLeu 446  
 QY 1529 GGGGACACTGGAGGTGTAGCAGTACAGCAGCCTGTATGACACCTCTCTCTGCGAGGCCCT 1588  
 Db 447 GlyValLeuAsnLeuAlaIleValValProGlnMetIleValSerValLeuAlaGlyPro 466  
 QY 1589 AAGCCTGGA-----GCTCCCTTCCCTATATGACACAGCTGGGGTCT 1627  
 Db 467 PheAspSerLeuPheGlyGlyGlyAsnLeuProAlaPheValValGlyAlaIleSerAla 486  
 QY 1628 GGAGGAGTGGCTGTCTCCCA-----CCTCCACCGCGCTCTCTGCGGGGCTCT 1675  
 Db 487 AlaIleSerGlyValLeuAlaIleValLeuLeuProLysProSerLysAspAlaAlaSer 506  
 QY 1676 GCCTGTGTGTCTCTCC 1690  
 Db 507 LysLeuSerLeuSer 511  
 RESULT 8  
 CGB07S  
 collagen alpha 1(III) chain - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 04-Dec-1986 #sequence-revision 04-Dec-1986 #text-change 07-May-1999  
 C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
 R:Fietzek, P.P.; Allmann, H.; Rautenberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
 A:Title: The covalent structure of calf skin type III collagen. I. The amino acid seq  
 A:Reference number: A02862; MUID:80026026; PMID:488906  
 A:Accession: A02862  
 A:Molecule type: protein  
 A:Residues: 1-242 <FIE>  
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
 A:Title: The covalent structure of calf skin type III collagen. II. The amino acid se  
 A:Reference number: A38001; MUID:80026027; PMID:488907  
 A:Accession: A38001  
 A:Molecule type: protein  
 A:Residues: 243-422 <DEW1>  
 R:Bentz, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
 A:Title: The covalent structure of calf skin type III collagen. III. The amino acid s  
 A:Reference number: A38002; MUID:80026028; PMID:488908  
 A:Accession: A38002  
 A:Molecule type: protein  
 A:Residues: 423-571 <BEN>  
 R:Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
 A:Title: The covalent structure of calf skin type III collagen. IV. The amino acid se  
 A:Reference number: A38003; MUID:80026029; PMID:488909  
 A:Accession: A38003  
 A:Molecule type: protein  
 A:Residues: 572-808 <LAN>  
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
 A:Title: The covalent structure of calf skin type III collagen. V. The amino acid seq  
 A:Reference number: A38004; MUID:80026030; PMID:488910  
 A:Accession: A38004  
 A:Molecule type: protein  
 A:Residues: 809-947 <DEW2>  
 R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
 A:Title: The covalent structure of calf skin type III collagen. VI. The amino acid se  
 A:Reference number: A38005; MUID:80026031; PMID:488911  
 A:Accession: A38005  
 A:Molecule type: protein  
 A:Residues: 948-1049 <ALL>



D	b	408	----	-----GlyProProGlySerProGlyP	415
Q	y	1492	GTAGAGGAGGCCAGTGTGTT	-----	1473
D	b	415	roArgGlyGlnProGlyValMetGlyPheProGlyProLysGlyAsnAspGlyAlaProG	435	
Q	y	1472	--AGGGCAGGATGTGCAGGGCTGAGAAGTGAACCCGGTGAGGCGGCTGAAGCTGTCA	1415	
D	b	435	lyLysAsnGlyGlu-ArgGlyGlyProGlyGlyProGlyProGlnGlyProAlaGlyLys	454	
Q	y	1414	CACGGCCACACT-----GTGGACAGGCATGTGGCACCGGCACGCCACAGGAAAGCTGC	1361	
D	b	455	AsnGlyGluThrGlyProGlnGlyProProGlyProThrGlyProSerGlyAspLysGly	474	
Q	y	1360	CACACTGCCCAATAGACTGCTCAGTGCAGATCCGAATCGCTGCACCA-----GCCGGTCCAT	1307	
D	b	475	AspThrGlyProProGlyProGlnGlyLeuGlnGlyLeu--ProGlyThrSerGlyProp	494	
Q	y	1306	GACCAGAGAGA-----AGACCAGGAGATGGCCACTGTGCAGAACACAGCCCCAG	1259	
D	b	494	roGlyGluAsnGlyLysProGlyGluProGlyProLysGlyGluAlaGly--AlaProG	513	
Q	y	1258	GCTGCCCATCCGAAGCGCTTCATCATAGTGTCTCCGGCGCTCGGTGCCCGCTCAGCTCT	1199	
D	b	513	lylleProGlyGlyLysGlyAspSerGlyAlaProGlyGluArgGlyPro-----	529	
Q	y	1198	GGGCACGCCCTGTTACAGCC-----CCTCGCCACGAAATCCGTGTAAACACGCGTAA	1145	
D	b	530	-----ProGlyAlaGlyGlyProProGlyProArgGly-----	540	
Q	y	1144	GGTCATGAGTGCATCCAGCTGCACAGCTCAGCCACGAGAAGCGCGCAGGTGCGGG	1085	
D	b	541	-----GlyAlaGlyProProGlyProGlyProGluGlyGlyLysGlyA	553	
Q	y	1084	CATCGGSC-----AGCACAGCTGGTCAGCGCGGGGAAGCAGGCGGCCAG	1040	
D	b	553	laAlaGlyProProGlyProProGlySerAlaGlyThrProGlyLeuGlnGlyMetProG	573	
Q	y	1039	GTTC-----GGAAAGCCAAAGCGGCCCGC-----ATGGACAGCA	1004	
D	b	573	lyGluArgGlyGlyProGlyGlyProGlyProLysGlyAspLysGlyGluProGlySerS	593	
Q	y	1003	GTGGGGCGACA-----AGGAGGGGGCGCACAGCCCTTCTGCTGG	965	
D	b	593	erGlyValAspGlyAlaProGlyLysAspGlyProArgGlyProThr-----	608	
Q	y	964	CTCGTGGGCGCAGCGCTCCCTCCACAGCCACGACGTGTGGCTGCTACGACGTGAG	905	
D	b	609	-----GlyProIleGlyProProGlyProAla-----	618	
Q	y	904	GAAGATGAGGTTGAGCAGGCCAAAGAGGCACTCTCTGTGGTGGCCAGGT-----	848	
D	b	618	lyGlnProGlyAspLysGlyGluSerGlyAla--ProGlyValProGlyIleAlaGlyP	637	
Q	y	847	CAGGGCACTGGTCTCCAGTCAATGGCAGCGAGGTAGTCCACGAGCGCCCCCAAGACT	788	
D	b	637	ro-----ArgGlyGlyProGlyGlu-----	643	
Q	y	787	GATCATGAAGGCATAGACAGTAGTGGCTGGCCACAGTGGT-----CCCGGAA	734	
D	b	644	-----ArgGlyGluGlnGlyProProGlyProAlaGlyPheProGlyAlaProGly	661	
Q	y	733	GAGGTCAAGACAGCAGGCGCTCCAGTGGAGTGAAGCACACTGTGCCACAGAGTCCAGCAG	674	
D	b	661	lnAsnGlyGluProGlyAlaLysGlyGluArgGlyAlaProGlyGluLysGlyGluGlyC	681	
Q	y	673	CCCCAGCCCGATGAGCATGTGCCAGCTCCAGGG-----GCCTGGATCCGGGCA	623	
D	b	681	lyPro--ProGlyAlaAlaGlyProAlaGlyGlySerGlyProAlaGlyProProGlyP	700	
Q	y	622	CAGCAGCCCTGCTACCGACCGGCCCTTGGATGAGAAAGAGGCTCAGCAGGATGCCCAA	563	
D	b	700	roGlnGlyLysGly-----GluArgGlySerProGlyGlyProG	714	

562 GGACAGTCCCCAGATGAAGGCCGGC-----GGCGGCCATAGCGTCCACGCCAGTGTGC 509

714 lYala-----AlaGlyPheProGlyGlyArgGly----- 723

508 ACTGGCTGAGCCTAGGACGGGACACAGACAGCAGGCCCGCCAGCACTGGACCAANTGCCAGCAC 449

724 -----ProProGlyPro----- 728

448 CATGTCATGAACCTTCTCTACCCCA-----CTTCCAGCAGCAGAGCGG 401

728 roGlySerAsnGlyAsnProGlyProProGlySerSerGlyAlaProGlyLysAspGly- 747

400 CACATAGGTGATGCTCGGGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTGACCA 341

748 -----ProProGlyProProGlySerAsn-----GlyAlaProG 759

340 CAAGAGCTGGCTTTTCGGTGCCGACGAGCG- - - - -GGCTCAC 302

759 lySerProGlyIleSerGlyProlySerGlyAspSerGlyProProGlyGluArgGlyAlaP 779

301 CCACAGCCTCTGSGACCATAGTGGG---CCAGCGGGTAGGCTCAGGGGGCGGTTTCAGGC 245

779 roGlyProGlnGlyPro-ProGlyAlaProGlyProLeuGlyIleAlaGlyLeuThrgly 798

244 ACTCCAGAACTGCTGCTCTCGGCTCTGCTCCAGAGCTGGCGCTCTCTCT- - - - - 194

799 AlaArgGlyLeuAlaGlyProProGlyMetProGlyAlaArgGlySerProGlyProGln 818

193 -----CCTTGCTGCCGCCAACTGCCTAGGAATCAG 164

819 GlyIleLysGlyGluAsnGlyLysProGlyProSerGlyGlnAsnGlyGluArgGlyPro 838

163 CCAGGCGCCCATTTCTGCCAGCCCTTGGT-----GCCGTCCAGCTTCTCAG 116

839 ProGly-----ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGlu 853

115 CCCATGCTC-----AACACTGCTGCTGTGGGCACTCTCACTGGGACAGCTCTCATCA 62

854 ProGlyArgAspGlyAsn-ProGlySerAspGlyLeuProGlyArgAspGlyAlaProG 873

61 CTCAGATCTCGCCGCGGCGCGCTGTCTACCCGGAGGCC 22

873 yAlaLysGlyAspArgGlyGluAsnGlySerProGlyAla 886

RESULT 9

CGHULS

collagen alpha 1(I) chain precursor - human

N:Alternate names: procollagen alpha 1(I) chain

C:Species: Homo sapiens (man)

C:Date: 12-Aug-1991 #sequence\_revision 04-Oct-1996 #text\_change 31-Dec-2000

C:Accession: 160114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S269; A29439, M3; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, F.; R.D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, F.; Gene 67, 105-115, 1988

A:Title: Complete nucleotide sequence of the region encompassing the first twenty-fi

A:Reference number: 160114; MUID:88329734; PMID:2843432

A:Accession: 160114

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-369, 'L', 371-589 <DAL>

A:Cross-references: GB:M20789; NID:g179593; PIDN:AAB59373.1; PID:g179594

P:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; P

Biochem. J. 253, 919-922, 1988

A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of hum

A:Reference number: S01143; MUID:89025644; PMID:3178743

A:Accession: S01143

A:Molecule type: mRNA

A:Residues: 1-472 <TRO>

A:Cross-references: EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID:g30016; GB:M36546;

A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, Nature 310, 337-340, 1984

A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of  
A:Reference number: A93335; MUID:84270697; PMID:6462220  
A:Accession: A93335  
A:Molecule type: DNA  
A:Residues: 1-58, Q', 60-181 <CHU>  
A:Cross-references: EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID:g35658  
R:Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.  
J. Biol. Chem. 262, 15151-15157, 1987  
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh  
A:Reference number: I55254; MUID:88033098; PMID:2822714  
A:Accession: I55254  
A:Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-45 <ROS>  
A:Cross-references: GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:g180388  
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.  
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987  
A:Title: Regulatory elements in the first intron contribute to transcriptional control c  
A:Reference number: A39943; MUID:88097389; PMID:3480516  
A:Accession: A39943  
A:Molecule type: DNA  
A:Residues: 1-34 <BOR>  
A:Cross-references: GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:g553238  
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.  
J. Biol. Chem. 260, 2315-2320, 1985  
A:Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s  
A:Reference number: I55237; MUID:85130970; PMID:2857713  
A:Accession: I55237  
A:Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-34 <CH2>  
A:Cross-references: GB:M10627; NID:g180383; PIDN:AAA51992.1; PID:g553226  
R:Wirtz, M.K.; Keene, D.R.; Horii, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollis  
J. Biol. Chem. 265, 6312-6317, 1990  
A:Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina  
tome, type VII.  
A:Reference number: A35233; MUID:90202908; PMID:2318855  
A:Accession: A35233  
A:Molecule type: protein  
A:Residues: 33-52 <WIR>  
A:Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved  
R:Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.  
EMBO J. 8, 1705-1710, 1989  
A:Title: A base substitution in the exon of a collagen gene causes alternative splicing  
A:Reference number: S09400; MUID:89356643; PMID:2767050  
A:Accession: S09400  
A:Molecule type: mRNA  
A:Residues: 156-183 <WEI>  
R:Click, E.M.; Bornstein, P.  
Biochemistry 9, 4699-4706, 1970  
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1  
A:Reference number: A90567; MUID:71038625; PMID:5529814  
A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5  
A:Accession: B90567  
A:Molecule type: protein  
A:Residues: 162-198, Z', 200-201, Z', 203-206, Z', 208-209, Z', 211-228, B', 230, BB', 233, Z'  
A:Experimental source: skin  
A:Note: evidence for 170-allysine  
R:Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, H.  
Eur. J. Biochem. 192, 153-159, 1990  
A:Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle  
A:Reference number: S11372; MUID:90382436; PMID:2169412  
A:Accession: S11372  
A:Molecule type: protein  
A:Residues: 175-187; 274-287, P', 289 <BAE>  
A:Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion  
R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez  
J. Biol. Chem. 266, 21827-21832, 1991  
A:Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain  
operative melting of intact type I collagen.  
A:Reference number: I55342; MUID:92042092; PMID:1718984  
A:Accession: I55342  
A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 258-268; 1347-1357 <DEA>  
A:Cross-references: GB:S67495; NID:g239007; PIDN:AAB20350.1; PID:g239008  
A:Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg rep  
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.  
J. Biol. Chem. 245, 5042-5048, 1970  
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagen  
A:Reference number: A92069; MUID:71001508; PMID:4319110  
A:Accession: A92069  
A:Molecule type: protein  
A:Residues: 263-268 <MOR>  
A:Experimental source: skin  
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine  
R:Labhard, M.E.; Hollister, D.W.  
Matrix 10, 124-130, 1990  
A:Title: Segmental amplification of the entire helical and telopeptide regions of the  
A:Reference number: S15989; MUID:90326017; PMID:2374517  
A:Accession: S15989  
A:Molecule type: mRNA  
A:Residues: 281-302; 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <LAB>  
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries,  
Connect. Tissue Res. 29, 1-11, 1993  
A:Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain  
A:Reference number: I52905; MUID:93339042; PMID:8339541  
A:Accession: I52905  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 342-352, C', 354-359 <WID>  
A:Cross-references: GB:S64717; NID:g408195; PIDN:AAB27677.1; PID:g408196  
A:Note: mutant sequence from patient with osteogenesis imperfecta  
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.  
Biochemistry 22, 5213-5223, 1983  
A:Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalph  
A:Reference number: A90476; MUID:84080385; PMID:6689127  
A:Accession: A90476  
A:Molecule type: mRNA  
A:Residues: 425-1250, X', 1252-1328, S', 1330-1390, X', 1392-1464 <BER>  
A:Cross-references: GB:K01228; NID:g80391; PIDN:AAA51995.1; PID:g80392  
A:Note: sequence partially completed for missing nucleotides by A29439  
R:Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.  
J. Biol. Chem. 260, 691-694, 1985  
A:Title: Multixon deletion in an osteogenesis imperfecta variant with increased type  
A:Reference number: A22161; MUID:85104934; PMID:2981843  
A:Accession: A22161  
A:Molecule type: DNA  
A:Residues: 472-594, R', 596-607 <CH3>  
A:Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PIDN:AAA51847.1;  
A:Note: the authors translated the codon CGT for residue 595 as Pro  
R:Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.  
Am. J. Hum. Genet. 46, 1034-1040, 1990  
A:Title: Variable expression of osteogenesis imperfecta in a nuclear family is explai  
A:Reference number: A35336; MUID:90252792; PMID:2339700  
A:Accession: A35336  
A:Molecule type: mRNA  
A:Residues: 710-720, E', 722-737, E', 739-745 <WAL>  
A:Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu  
R:Forlino, A.; Zollezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mot  
Hum. Mol. Genet. 3, 2201-2206, 1994  
A:Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in th  
A:Reference number: I54365; MUID:95187161; PMID:7881420  
A:Accession: I54365  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 746-766, S', 768-781 <FOR>  
A:Cross-references: GB:I47667; NID:g1009093; PIDN:AAB59576.1; PID:g1009094  
R:Chessler, S.D.; Wallis, G.A.; Byers, P.H.  
J. Biol. Chem. 268, 18218-18225, 1993  
A:Title: Mutations in the carboxy-terminal propeptide of the pro alpha 1(I) chain of  
A:Reference number: A47426; MUID:93352646; PMID:8349697  
A:Accession: A47426  
A:Molecule type: mRNA  
A:Residues: 1179-1276, H', 1278-1336, 1339-1387, R', 1389-1464 <CHE>

**Alignment Scores:**

US-09-759-143-110 (1-3410) x CGHUIS (1-1464)

QY 971 CTGCTGGCTCGTGGGCGCCAGCG-----CTGCTCTCTCAGCCACCA 930  
Db |||||  
QY 557 -----GlyProAlaGlyGlnAspGlyArgProGlyProGlyProPro 571  
Db |||||  
QY 929 -----GCAGTGTGGCTG 918  
Db |||||  
QY 572 GlyAlaArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAla----- 589  
QY 917 CTACCGAGGTGAGGAAGATGAGGTGAGCGAGCGCAAGAGGCACT-----CCTCTGGG 864  
Db |||||  
QY 590 -----GlyGluProGlyLysAlaGlyGluArgGlyValProGlyProProGly 605  
Db |||||  
QY 863 TGCCCGAGT-----AGC 852  
Db |||||  
QY 606 AlaValGlyProAlaGlyLysAspGlyGluAlaGlnGlyProGlyProAla 625  
QY 851 GGGCCAGGCGCACTGTGTCCAGTCAATGGCAGCGAGGA----- 813  
Db |||||  
QY 626 GlyProAlaGlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeu 645  
QY 812 ---GATAGCCAGGCGCCAGAGCTGATCATGAGGCATAGACAGTAGCCCTGGC 756  
Db |||||  
QY 646 ProGlyProAlaGlyProGlyGlyAlaGlyLysProGlyGluGlnGlyValProGly 665  
QY 755 GACAGTGT-----CGGGTCCC-----GGAAGAGTGCAGACAGCGGCTCCAGTGGAG 705  
Db |||||  
QY 666 AspLeuGlyAlaProGlyProSerGlyAlaArgGlyGluArg-----GlyPheProGlyGlu 684  
QY 704 TGAAGCACCTGCGCACAGAGTCCAGACGCCCGCCAGGATGAGCAGTGCCTCA----- 648  
Db |||||  
QY 685 ArgGlyValGlnGlyProGlyProAlaGlyProArgGlyAlaAsnGlyAlaProGly 704  
QY 647 -----CCTCAGGGGCTGGATCGGGGCGACAGCAGCC 615  
Db |||||  
QY 705 AsnAspGlyAlaLysGlyAspAlaGlyAlaProGlyAla-----ProGlySerGlnGly 722  
QY 614 CTGTAGCAGCGCGCCCTTGGATGAGAAAGAGGCTCAGCAGGATGCCAAGCAGCAGTG 555  
Db |||||  
QY 723 AlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGly----- 740  
QY 554 CCCAGATGAGGCGCGGGCGGCCATAGCTCCAGCGCAGTGTCTGCTGCTGAGCCTA 495  
Db |||||  
QY 741 ProLysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGly-----SerPro 757  
QY 494 GGAGCG---GGACACAGA-----CCAGGCGCCAGCAGTGGAC 462  
Db |||||  
QY 758 GlyLysAspGlyValArgGlyLeuThrGlyProIleGlyProProGlyProAla----- 775  
QY 461 CAATGCCCA----- 453  
QY 776 GlyAlaProGlyAspLysGlyGluSerGlyProSerGlyProAlaGlyProThrGlyAla 795  
QY 452 -----GCACCATGTCATGAATCTCTCTACCCCGCAGCAGGCGGTCCACCCACCCCTCT 399  
Db |||||  
QY 796 ArgGlyAlaProGlyAspArgGlyGluProGlyProGlyProAlaGlyPheAlaGly 815  
QY 398 CATAGTGTGCTGCTGGCGCAACACACACCTCCAGGC-----CAAGGTTAGCA 351  
Db |||||  
QY 816 -----ProGlyAlaAspGlyGlnProGlyAlaLys 826  
QY 350 GGTGTACACAGACAGCTGGCTTCCGTGCGGCGCAGCAGGCGGTCCACCCACCCCTCT 291  
Db |||||  
QY 827 GlyGluProGlyAspAlaGlyAlaLysGlyAspAla-----GlyProProGlyProAla 844  
QY 290 GGACCATGTGGCCA-----GCGGGTAGGCTCAGGGGCGGTTCAGGCAC 243  
Db |||||  
QY 845 GlyPro-AlaGlyProProGlyProIleGlyAsnValGlyAlaProGlyAlaLysGlyAla 864  
QY 242 TCCAGAAGTCTGCTGCTGCTGCTGCT-----CCAGAAGCTGGCGGCTCTCTCTCTCT 189  
Db |||||  
QY 864 a-----ArgGlySerAlaGlyProProGlyAlaThrGlyPheProGlyAla 879

QY 188 CTCCGCCCAACTGCCTAGGAATCAGCCAGCGCCCATTTCTGCCAGCCCTTTGGTGGCGG 129  
Db |||||  
QY 879 aAlaGlyArgValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGly 899  
QY 128 TCAGGTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGG-----GCACCTCAGTGGG 75  
Db |||||  
QY 899 yProAlaGlyLysGlu-----GlyGlyLysGlyProArg-GlyG 912  
QY 74 ACAGTCTCATCACTCACTCAGATCCTGTGGCGCA-----G 45  
Db |||||  
QY 912 luhrglyProAlaGlyArgProGlyGluValGlyProProGlyProProGlyProAlaG 932  
QY 44 GCGCGCGGCTGTCAACCGGAGCC 22  
Db |||||  
QY 932 lyGluLysGlySerProGlyAla 939

RESULT 10  
CGCHIS  
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)  
C:Species: Gallus gallus (chicken)  
C:Date: 12-Aug-1981 #sequence\_revision 06-Jul-1982 #text\_change 31-Mar-2000  
C:Accession: A90458; A90181; A02857  
R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, Biochemistry 21, 2048-2055, 1982  
A:Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete pr  
A:Reference number: A90458; MUID:82231995; PMID:7093229  
A:Accession: A90458  
A:Molecule type: protein  
A:Residues: 1-1036 <HIG>  
A:Experimental source: skin  
A:Note: this is the latest in a series of papers from these workers elucidating the s  
R:Eyre, D.R.; Glimcher, M.J.  
Biochem. Biophys. Res. Commun. 48, 720-726, 1972  
A:Title: Evidence for a previously undetected sequence at the carboxyterminus of the  
A:Reference number: A90181; MUID:72243016; PMID:5047697  
A:Accession: A90181  
A:Molecule type: protein  
A:Residues: 1037-1042 <EYR>  
A:Experimental source: skin  
A:Note: residues 1037-1042 above correspond to the carboxyl end of the protein  
C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in s  
C:Comment: Most of the prolines at the third position of the tripeptide repeating uni  
C:Comment: pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trime  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Alignment Scores:  
Pred. No.: 4,1e-13 Length: 1042  
Score: 342.00 Matches: 270  
Percent Similarity: 32.15% Conservative: 48  
Best Local Similarity: 27.30% Mismatches: 385  
Query Match: 5.49% Indels: 286  
DB: 1 Gaps: 52

US-09-759-143-110 (1-3410) x CGCHIS (1-1042)

QY 2578 ATGCACACAGAGACCAACAGCCCATCTGTATAAAGGTAAAGGGGGTGGATCAG 2519  
Db |||||  
QY 55 MetGlyProArgGlyProAlaGly--ProProGlyLysAsnGlyAspGlyGluAlaG 74  
QY 2518 CAAAAGACACAGTGTGGGCTGAGGGACCT-----GGTTCTTGTGTGTGC 2471  
Db |||||  
QY 74 lyLysProGlyArgProGlyGlnArgGlyProProGlyProGlnGlyAlaArgGlyLeuP 94  
QY 2470 CC---CTCAGGACTCTTCCCTACAAAATAGTCATATG-----T 2435  
Db |||||  
QY 94 roGlyThrAlaGlyLeuProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyA 114  
QY 2434 TCAATCCCATGGAGAGTGTTCATCTCAGAAACTCCCATGCAAGAGCTACATTAACG 2375  
Db |||||  
QY 114 laLysGlyGlnProGlyProAlaGlyProLysGlyGluProGlySerProGlyGluAsnG 134







Db 756 rgProGlyGluProGlyProAlaGlyProProGlyProProGlyGluLysGlySerProG 776  
QY 520 AGCCAGTGGTCACTGGCTGAGCCCTAGGAGGGGACACAGACACAGGCCCA----- 471  
Db 776 lyAlaAspGly-----ProIleGlyAlaProGlyThrProGlyProGlnGlyIleA 793  
QY 470 -----GCACGTGGACCAAT 458  
Db 793 laGlyGlnArgGlyValValGlyLeuProGlyGlnArgGlyGluArgGlyPheProGlyL 813  
QY 457 GCCCAGACCATGTCATGAATCTCTCTACCCCCCTTCCAGCAGCAGAGCGGCAC 398  
Db 813 euProGlyProSerGly-----GluProGlyLysGlnGlyProSerGlyAlaSerGly 831  
QY 397 ATAGTGATGCTCGCGGCAACACACACTCCAGGCGCAAGGTAGCAGGTGTGACCAACAA 338  
Db 831 luArgGlyProProGlyProMetGlyProProGly-----LeuAlaGlyProProGly 849  
QY 337 GAGCTGGGCTTCCCGTGCC---GCACGAGGCGGCTCACCCACAGCCTCTCGACCATAGT 281  
Db 849 luAlaGlyArgGluGlyAlaProGlyAlaGluGlyAlaProGlyArgAspGlyAla-Ala 868  
QY 280 GGGCCA---GGCGGTAGGCTCAGGGGCGCTTCAGCACTCCAGAACTGCTTCGCTC 224  
Db 869 GlyProLysGlyAspArgGlyGluThrGlyProAlaGlyProPro----- 883  
QY 223 GGCCTGCTCCAGAGCTCGCGCTCTCTCT-----CCTTGCTGCGGCCAATG 176  
Db 884 ---GlyAlaProGlyAlaProGlyAlaProGlyProAlaGlyLysAsnGly 902  
QY 175 CCTAGGAATCAGCAGCGGCCCTTCTGCGCAGCCCTTGGGCGGCTCCAGCTTCTCAG 116  
Db 903 AspArgGlyGluThrGlyPro-----AlaGlyProAlaGlyPro 915  
QY 115 CCC 113  
Db 916 Pro 916

RESULT 11  
T12198  
sucrose transport protein - fava bean  
C:Species: Vicia faba (fava bean)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 20-Jun-2000  
C:Accession: T12198  
R:Weber, H.; Borisjuk, L.; Helm, U.; Sauer, N.; Wobus, U.  
Plant Cell 9, 895-908, 1997  
A:Title: A role for sugar transporters during seed development: molecular characterization of the sucrose transporter gene from Vicia faba  
A:Reference number: 217451; MUID:97355984; PMID:9712465  
A:Accession: T12198  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-523 <WEB>  
A:Cross-references: EMBL:293774  
C:Genetics:  
C:Gene: sut  
C:Superfamily: common tobacco sucrose transport protein

Alignment Scores:  
Pred No.: 7,67e-13 Length: 523  
Score: 337.50 Matches: 134  
Percent Similarity: 40.38% Conservative: 76  
Best Local Similarity: 25.77% Mismatches: 185  
Query Match: 5.26% Indels: 125  
DB: 2 Gaps: 20

US-09-759-143-110 (1-3410) x T12198 (1-523)

QY 332 CAGCTCTGCTGTCACCTGTAACCTTTGGCGCTGGAGGTGTGTTGGCCGAGGCATC 391  
Db 36 LysileMetValValAlaSerIleAlaAlaGlyValGlnPheGlyTrpAlaLeuGlnLeu 55  
QY 392 ACCTATGCGCGCTCTGCTGCTGGAAGTGGGGGTAGAGGAGAGGATTCATGACCATGGTG 451

Db 56 SerLeuLeuThrProTyrrValGlnLeuLeuGlyIleHisHisThrTrpAlaAlaTyrrIle 75  
QY 452 CTGGGCATTTGGTCCAGCTGGGCTGGTCTGTGTCCCGCTCTAGCTACGCCAGTGCAC 511  
Db 76 TrpLeuGlyGlyProIleSerGlyMetLeuValGlnProIleValGlyTyrrHisSerAsp 95  
QY 512 CACTGGCTGGAGCTATGAGCCCGCGGCGCTTCTATCTGGGCACTGCTCTGGGCATC 571  
Db 96 ArgCysThrSerArgPheGlyArgArgProPheIleAlaAlaGlySerIleAlaVal 115  
QY 572 CTGCTGAGCTCTTCTCATCCCAAGGCGGCTGGCTAGCAGGGCTG----- 619  
Db 116 AlaIleAlaValPheLeuIle-----GlyTyrrAlaAlaAspLeuGlyHisSerPhe 132  
QY 620 -----CTGTCGCCGGATCCAGGCGGCTGGAGCTGGCACTGCATCTATCTGCTGGCGTG 670  
Db 133 GlyAspSerLeuAspGlnLysValArgProArgAlaIleGlyIlePheValGlyPhe 152  
QY 671 GGCTGCTGAGCTCTGTGGCCAGGTGTGTCTACCTCCAGCTGGAGCGCTGCTCTCTGAC 730  
Db 153 TrpIleLeuAspValAlaAsnAsnMetLeuGlnGlyProCysArgAlaLeuLeuGlyAsp 172  
QY 731 CTCTTCCGG---GACCGGACCACTGTGCGCCAGGCGCTCTCTCTATGCTTCTATGATC 787  
Db 173 LeuCysAlaGlyAsnGlnArgLysThrArgAsnAlaAsnAlaPheSerPheMet 192  
QY 788 AGTCTGGGGGCTGCTGGGCTACCTCTCTGCGCCAGGCGCTCTCTCTATGCTTCTATGATC 829  
Db 193 AlaValGlnAsnValLeuGlyTyrrAlaAlaGlyAlaIleTyrrSerLysLeuTyrrHisValPhe 212  
QY 830 -----TGGGACACCACTGCGCTGGCGGCTCTCTCTGCGCCAGGAGGAGTGCCTTT 883  
Db 213 ProPheThrLysThrLysAlaCysAsnValTyrrCysAlaAsnLeuLysSerCysPhePhe 232  
QY 884 GGCTGCTGCTACCTCTCTCTACCTGCTGAGCAGCAGCTGCTGGTG---GCTGAG 940  
Db 233 -----LeuSerIleAlaLeuLeuThrValLeuAlaThrSerAlaLeuIleTyrrValLys 250  
QY 941 GAGCAGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 988  
Db 251 GluThrAlaLeuThrProGluLysThrValValThrValThrGluAspGlyGlySerSerGly 270  
QY 989 TCCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1042  
Db 271 GlyMet-----ProCysPheGlyGlnLeuSerGlyAlaPheLysGluLeu 285  
QY 1043 GGCCTGCTGCTTCCCGGCTGCACAGCTGTGCTGCGCATGCCCGGCGGCGGCGGCGGCGGCGG 1102  
Db 286 -----LysArgProMetTrpIle 291  
QY 1103 CTCCTGCTGGCTGAGCTGTGCGAGCTGGATGGCACTCATGACCTTCACGCTCTTTTACAGC 1162  
Db 292 LeuLeuLeuValThrCysLeuAsnTrpIleAlaTrpPheProPheLeuLeuPheAspThr 311  
QY 1163 GATTTCGTGGCGAGGCGGCTGTACAGGCGGCTGCCAGAGCTGAGCGGCGGCGGCGGCGGCGGCGG 1222  
Db 312 AspTrpMetGlyLysGluValTyrr-----GlyGlyThrValGly 324  
QY 1223 CGGAGACAC---TATGATGAAGGCTTCGGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1279  
Db 325 GluGlyHisAlaTyrrAspMetGlyValArgGluGlyAlaLeuGlyLeuMetLeuAsnSer 344  
QY 1280 GCCATCTCCCTGCTCTCTGCTGATGAGCGGCTGTGTCAGCGATTCGGGCGGCGGCGGCGGCGGCGG 1339  
Db 345 ValValLeuGlyAlaThrSerLeuGlyValAspIleLeuAlaArgGlyValGly---Gly 363  
QY 1340 GCAGCTCTATTTGGCGAGTGTGGCAGCTTTCCCTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1399  
Db 364 ValLysArgLeuTrpGlyIleValAsnPheLeuLeu-----AlaIleCysLeuGly 380  
QY 1400 CACAGTGGCGGCTGGTG----- 1417  
Db 381 LeuThrValLeuValThrLysLeuAlaGlnHisSerArgGlnTyrrAlaProGlyThrGly 400

QY 1418 -----ACAGTTCACGCGCCCTCACCGGGTTC 1444  
 Db 401 AlalaLeuGlyAspProLeuProSerGluGlyIleLysAlaGlyAlaLeuThrLeuPhe 420  
 QY 1445 ACCTTTCACGCGCCCTGCAGATC-----CTGCGCTACACACTGGGCTCCCTC 1489  
 Db 421 SerValLeuGlyValProLeuAlaIleThrTyrSerIleProPheAlaLeuAlaSerIle 440  
 QY 1490 TACCACCGGAGAGCAGGTTCCTGCCCAATACCGAGGGGACACTGGAGGTGTAGC 1549  
 Db 441 Phe-----SerSerThrSerGlyAlaGly 448  
 QY 1550 AGTGAGGACAGCCTG-----ATGACACG 1573  
 Db 449 GlnGlyLeuSerLeuGlyValLeuAsnLeuAlaIleValIleProGlnMetPheValSer 468  
 QY 1574 TTCTCGCCAGCGCCCTAAGCCTGGA-----GCTCCCTTCCCTAAT 1612  
 Db 469 ValLeuSerGlyProTrpAspAlaLeuPheGlyGlyAsnLeuProAlaPheValVal 488  
 QY 1613 GGACACGTGGTCTGGAGGACAGTGC-----CTGCTCCACCTCCACCC 1657  
 Db 489 GlyAlaValAlaAlaLeuAlaSerGlyIleLeuSerIleLeuLeuProSerProPro 508

## RESULT 12

I50694  
 collagen alpha 1(III) chain - chicken (fragment)  
 C:Species: Gallus gallus (chicken)  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 13-Aug-1999  
 C:Accession: I50694  
 R:Nah, H.D.; Niu, Z.; Adams, S.L.  
 J. Biol. Chem. 269, 16443-16448, 1994  
 A:Title: An alternative transcript of the chick type III collagen gene that does not end  
 A:Reference number: A54041; MUID:94266842; PMID:8206952  
 A:Accession: I50694  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-886 <N>A>  
 A:Cross-references: EMBL:U07973; NID:g520454; PIDN:AAA83407.1; PID:g537432  
 C:Genetics:  
 A:Gene: COL3A1  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 F:30-90/Domain: von Willebrand factor type C repeat homology <WVC>

## Alignment Scores:

Pred. No.: 1,01e-12 Length: 886  
 Score: 335.50 Matches: 270  
 Percent Similarity: 31.63% Conservative: 46  
 Best Local Similarity: 27.03% Mismatches: 336  
 Query Match: 5.39% Indels: 348  
 DB: 2 Gaps: 58

US-09-759-143-110 (1-3410) x I50694 (1-886)

QY 2479 GTGTGT---TGCCCTCAGGACTCTCCCTACAATAAGTCATATGTCATAATCCCATG 2423  
 Db 51 IleCysValCysAspSerGlySerValLeuCysAspAspIleIleCysAspAspGlnGlu 70  
 QY 2422 GAGGAGTGTTCATCTAGAAATCCCATGCAAGAGCTACATTAACGAAGCTGCAGTT 2363  
 Db 71 LeuAspCysProAsnProGluIleProLeuGlyGluCys-----CysProVal 86  
 QY 2362 AAGGGCTTAGAGTGGGAAACAGGCTGACTGAGTTTATTC-----GCTCCCAAAACCCCT 2306  
 Db 87 CysProGlnThrProGlnProThrGluLeuProTyrThrGlnGlyProLys----- 104  
 QY 2305 TCTCTAGGTGTCTCAACTAGGAGGTAGCTGTTAACCCCTGAGCCTGGGTAAATCCA--- 2249  
 Db 105 -----GlyAspProGly 108  
 QY 2248 ---CCTGCAGAGTCCCGCAGTCCAGTGCATGAGGACCTCTTGGCTCCCTGTATAGTC 2192  
 Db 403 ---ProGlyGlyProGlyProGlyGlyArgGlyLeuProGlyProGlyThrSer 421

Db 109 SerProGlySerProGlyArgThrGlyAlaProGlyProProGlyGlnPro----- 125  
 QY 2191 CAGACTGAAACCCCTTGAAGGCTCCAGTCAGCAGCAGCCTTAGAGACTGGGAGAGAGG 2132  
 Db 126 -----GlySerPro----- 128  
 QY 2131 AGAGGAGCAGCCCGCCAGTGTGACGTGTGACGTACGCACCTCAGCAGCAGCAGGTGGCAGC 2072  
 Db 129 ---GlyAlaProGlyProProGlyIleCysGlnSerCysProSerIleSerGlyGlySer 147  
 QY 2071 AGAGGAGCAGCATTC-----TTTGGCAGCAACAGAACTGGCGGCA 2030  
 Db 148 PheSerProGlnThrAspSerTyrAspValLysAlaGlySerValGlyMetGlyTyrPro 167  
 QY 2029 GCCCGGAG-----CCCATGGGCTAACAGAGCGGGGAGCTGGGA---CCC 1985  
 Db 168 ProGlnProIleSerGlyPheProGlyProProGlyProSerGlyProProGlyProPro 187  
 QY 1984 AGTGAGGAGCAGCCCTCCACCCCAATGTCTGGAAGTTTCTACGCTAGTATTTGGCCAA 1925  
 Db 188 GlyHisAlaGlyProPro-----GlySer----- 195  
 QY 1924 CTCGCTCTGTCAATACTACTCTGTGTAGCAAAGTAAATGGCAGCAGAC---CCAGGC 1869  
 Db 196 -----AsnGlyTyrGlnGlyProProGly 203  
 QY 1868 CTGCGGAGACACCATATAGCAGTGTGACAGTGTGCTGAGTGTGACATGTGAGCCATAA 1809  
 Db 204 GluProGlyGlnProGlyProSerGlyProProGlyProAlaGlyMetIleGlyPro--- 222  
 QY 1808 ACAGGATGGGCGCCACTGGCAGCAGCAGGAGGACACTCCAGGATGCGGAGGTCCAGGC 1749  
 Db 223 -----AlaGlyProProGlyLysAspGlyGlu-----ProGly 233  
 QY 1748 AGATGCCCGCGCGGAGAC-----CA 1728  
 Db 234 Arg---ProGly-ArgAsnGlyAspArgGlyIleProGlyLeuProGlyHisLysGlyHl 252  
 QY 1727 CCTTGGCTCGTGGGTCAACC-----CACCACCA 1698  
 Db 252 sProGlyMetProGlyMetProGlyMetLysGlyAlaArgGlyPheAspGlyLysAspGl 272  
 QY 1697 CAGCTACGAGACATCACAGCAGGCGCCGCGAG-----AGCGCGG 1656  
 Db 272 YAlaLysGlyAsp-SerGlyAlaProGlyProLysGlyGluAlaGlyGlnProGlyAlaA 292  
 QY 1655 GTGAGGTGGGAGCAGCCACTGCTCCAGCAGCAGCTGTCCATTAGGGAAG---GGAG 1599  
 Db 292 snGlySerProGlyGlnProGlyProGlyProThrGlyGluArgGlyArgProGlyA 312  
 QY 1598 CTCAGGCTTAGGCGCT-----GGCAGGAAGCTGTGCATCAGGCTGTCTCTCACTGC 1548  
 Db 312 snProGly---GlyProGlyAlaHisGlyLysAspGlyAlaProGlyThrAlaGlyProL 331  
 QY 1547 TAGCACCTCCAGTGTCCCTCGGTATTTGGGAGAGAACACTGCTCTCCCGGTGTAGA 1488  
 Db 331 euGlyProProGlyProPro-GlyThr---AlaGlyPheProGlySerProGlyPheLys 349  
 QY 1487 GGGAG-----CCAGTGTGTAGGCGAGGATCTGCAGGCTGAGAGGTGA 1443  
 Db 350 GlyGluAlaGlyProProGlyProAlaGlyAlaSerGlyAsnProGlyGluArg-GlyGl 369  
 QY 1442 ACCCGGTGAGGGCGGTGAAGCTGTACACGAGCCGACACTGTGGGAGCAGCATGTGGCAC 1383  
 Db 369 uProGlyProGlnGly--GlnAlaGlyProProGlyProGlnGlyProProGlyArgAla 388  
 QY 1382 CGCAGCCAGGAGAAAGCTGCCACACTGGCCCAAAATAGACTGTCTCGAGTCCGCAATCGCT 1323  
 Db 389 GlySerProGlyGlyLysGlyGluMetGlyProSerGlyIle----- 402  
 QY 1322 GCACCGCGGTTCATGACCA-----GAG 1299  
 Db 403 ---ProGlyGlyProGlyProGlyGlyArgGlyLeuProGlyProGlyThrSer 421



```

Db      76 ArgGlyThrProGlyIleThrGlyProLysGlyAspArgGlyGln-----ThrGlyThr 93
QY      2925 CCAGGCTGGCCCAAGAGAGAGAGGGTGGTTAGGAAGCGCTTGAGACCTGAAGCCGCCACC 2866
Db      94 ProGly-GluProGlyGluLysGlyGluArgGlyProPro-GlyProValGlyProGlnG 113
QY      2865 CTCTACCTCTCTTCAACACCCTAACCTTGGGTAAACAGCATTTGGAATTATCATTTGGGAT 2806
Db      113 LysLeuProGly-----ValAlaGly-----HisProGly- 122
QY      2805 GAGTAGAATTTCCAAAGTCTCTGGTTAGGCATTTTGGGGGGCCAGACCCCGAGAGAGAA 2746
Db      123 --ValGluGlyProGlu-----GlyProProGlyProAlaGlyArgA 136
QY      2745 GATTCTGGCAATGATCAGCCCAATGACCAAGTACAGCTATCTCAGGGACCTGATTGTGGGATC 2686
Db      136 rgGlyGluLysGlyGluProGlyArgPro-----GlyAspP 148
QY      2685 CCCCACCCTACCCA---AATATTAGACACCAACAGAAAAGCTAGCAATGATTCCCTT 2629
Db      148 roAlaValGlyProGlyGlyAlaGlyAlaLysGlyGluLysGlyAspAlaGlyLeuPro- 167
QY      2628 CTACTTTGTTAAATAAATAGTTAAATATTAAATGCCTGTCTCTGTGATGGCAACAG 2569
Db      167 ----- 167
QY      2568 AAGGACCAACAGGCACATCTGTATAAAGGTAAAGAGGGGGTGGTATCAGCAAAAGACA 2509
Db      168 --GlyProArgGlyAlaAlaGlyIleLysGlyGluGlnGly----- 180
QY      2508 GTGCTGTGGCTGAGGGACCTGGTCTTGTGTGTGGTCCCTCAGGACTCTTCCCTTACA 2449
Db      181 -----ProProGlyLeu----- 184
QY      2448 AATAAGTCATATTTCAATCCATCCATGGAGGAGTGTTCATCTCAAGAACTCCCATGCAAG 2389
Db      185 -----AlaLeuProGlyAspProGlyProLys----- 193
QY      2388 AGCTACATTAAACGAAGTGCAGGTTAAGGGCTTAGAGATGGGAAACCAAGTACTGAG 2329
Db      194 -----GlyAspProGlyAsp----- 198
QY      2328 TTATTACGCTCCCAAAACCCCTTCTCTAGTGTGTCTCAACTAGGAGGCTAGCTGTAA 2269
Db      199 -----ArgGlyPro-----IleGlyLeuThrGlyArgAlaGly-----P 210
QY      2268 CCCTGAGCTGGGTAATCCACCTGCAGAGTCCCGCATTCACAGTGCATGGAGCCCTCTG 2209
Db      210 roThrGlyAspSerGlyProProGlyGluLysGlyAspProGlyArgProGlyProProG 230
QY      2208 GCCTCCCTGTATAAGTCCAG-----ACTGAAACCCCTTGGAGGCC 2167
Db      230 ly---ProValGlySerArgGlyArgGlyGluValGlyGluLysGlyValGluGlyA 249
QY      2166 TC-----CAGTCAGCAGCCCTAGACTGGGGAGAGG-----AGAGGAGCGGCC- 2120
Db      249 snProGlyAspProGlyLeuProGlyLysAlaGlyGluArgGlyLeuArgGlyAlaProG 269
QY      2119 -----CAGCCCCCAGCTGTGCAGCTACCCACCTCAGCAGCAGGTCGCGCAGAGAGC 2065
Db      269 lyValArgGlyProAla-----GlyGluLysGlyAspGlnGlyAsp 283
QY      2064 CACATTACTTTGGAGCAACAAGAACTGGCGGCCAGCCCGGAGCCCATGGGCTAACA 2005
Db      283 roGlyGluAspGlyArgGlnGlySerProGlyProSer---GlyProLysGlyAspArg- 301
QY      2004 GGAGCGGGAGCTGGGACCCAGTAGGAGGCCCTCCACCCCAATGCTGGAAGTTTC 1945
Db      302 -----GlyGluProGlyPro-----ProGlyProProGlyArgLeuValGlyAlaGlyI 318
QY      1944 TACGCTGAGTATTGGCCCAAGTCGCTCTGTCAAACTACTACTCTGTAGCAAAATAATG 1895

```

---

```

Db      318 leGly-----SerArgAspLysG 324
QY      1884 GCGACCA-----GACCA----- 1872
Db      324 LysGluProGlyGlnGluGlyProArgGlyProLysGlyAspProGlyProGlyAlaAs 344
QY      1871 -----GGCTTGGCGCAGACACCATATAGCAGTGCAGACAGTGGCT 1832
Db      344 erGlyGluArgGlyIleGluGlyLeuArgGlyProProGlyProGlnGlyAspProGly 364
QY      1831 GAGCTGGGCAATGGAGCCCATAAACAGAGATGGGGCCACCTGGGACACAGAGGCACT 1772
Db      364 alArgGlyProAlaGlyAspLysGlyAspArgGlyProProGlyLeuAspGlyArgAsn- 383
QY      1771 ATCCAGGATGGCGAGTCCAGGCGATGCCCGGCCCGGAAACCCCTGGCTCGGTGG 1712
Db      384 -----GlyValAspGlyLysProGlyAlaProGlyPro----- 394
QY      1711 CTCACCCACCACACGTCAGGAGACATCAGACGAGAGGCCCCCGAGAGCGGGGTGG 1652
Db      395 -----ProGlyProHis--GlyAlaSerGlyLysAlaGlyAspProGlyArgAspGlyLe 412
QY      1651 AGTGGGAGCAGG-----CCACTG---CCTCCAGCACCCAGCTGTCCATTAGG 1607
Db      412 uProGlyLeuArgGlyGluHisGlyProAlaGlyProProGlyProGlyValProG 432
QY      1606 GAAG-----GGAGCTCCAGGCTTAGG----- 1585
Db      432 YLysThrGlyGluAspGlyLysProGlyLeuAsnGlyLysAsnGlyGluProGlyAspPr 452
QY      1584 -----CCTGGCAGAAAGCTGGTTCATCAGGCTGTCTCAGCTGTCTCAGTACAC 1541
Db      452 oGlyGluAspGlyArgLysGlyGluLys-GlyAspSerGlyAlaProGlyArgGluGlyP 472
QY      1540 TCCAGTGTCCCTCGGTATTGGGCGAGAAACACCTGTCTTCCCGGTGGTAGAGGAGGC 1481
Db      472 roAspGlyProLysGly---GluArgGlyAlaProGlyAspProGlyLeuArgGlyProp 491
QY      1480 CA-----GTGTGTAGGCGAGATCTGAGGGCTGAGAGGTGAACCGGTGAG 1433
Db      491 roGlyLeuProGlyGln-ValGlyProGlyGlnGlyPheProGlyValProGlyAsn 510
QY      1432 GCGCGCTGAAGCTCTCACCCAGCCACACTGTGGGACA----- 1395
Db      511 ThrGlyProLysGlyAspArgGlyGluThr--GlySerLysGlyGluGlnGlyLeuProG 530
QY      1394 -----GGCATGTGCACCGCAGCCAGCAGGAAAGCTGCCACACTGGCCAAATAGAC 1343
Db      530 lyGluArgGlyLeu-----ArgGlyGluProGlySerLeuProAsn---AlaGluArgL 547
QY      1342 TGCTCGAG-----TGCGAATCGCTGCA----- 1320
Db      547 euLeuGluThrAlaGlyIleLysValSerAlaLeuArgAspIleValIleGluThrTrpGly 567
QY      1319 --CCAGCGCGT-----CCATGACGAGAGAGACAGCAGGAGATGCGCAGCTGCAG 1271
Db      567 luSerSerGlySerPheLeuProValProGluArgArgProGlyProLysGlyAspProG 587
QY      1270 GAACA-----GCCCGAGCTGCCCATCCGAAACGCTTTCATC 1235
Db      587 lyGluArgGlyProGlyLysGluGlySerIleGlyPheProGlyGluArg----- 604
QY      1234 ATAGTGTCTCGGCGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCT 1175
Db      605 --GlyLeuLysGlyAspArgGlyAspProGlyProGlnGlyProGlyProGlyLeuAla 624
QY      1174 GCCCAGAAATCCGTGTAAACACAGCGTGAAGGTTCATGAGTGCATCCAGCTGCACAGCTC 1115
Db      624 lyGluArgGlyProGly-----ProProGlyLeuAlaGlyG 637
QY      1114 AGCCAGAGAGCGCGGCGAGGTGGGGCATCGCGCAGCAGCAGCTGTGTGAGCGGGG 1055
Db      637 luProGlyLysProGlyIlePro---GlyLeuProGly---ArgAlaGlySerAlaGlyG 655

```



Db 529 -----PheArgPheGly-----GlyGlyGly 535  
 QY 2774 TTTTGGGGGCCAG-----CCCGAGGAGAAGATTCTGGCAATGATC 2730  
 Db 536 AspAlaGlySerLysGlyProAsnValSerAlaGlnGluSerGlnAlaGlnAlaIleLeu 555  
 QY 2729 AGCCCAATGACAGCTATCTCAGGGGA-----CCT 2700  
 Db 556 GlnGlnAlaArgLeuAlaLeuArgGlyProAlaGlyProMetGlyLeuThrGlyArgPro 575  
 QY 2699 GATTGTTGGGGATCCCGCCCTACCCCAATATATAGACACACACAGAAAGCTAGCAA 2640  
 Db 576 GlyProMetGlyProGlySerGlyLeuLysGlyGluProGlyAspMetGlyPr 595  
 QY 2639 TGGATTCCCTTACTTTGTTAAATAAATAAGTTAAATATTTAAATGCCTGTCTCTCT 2580  
 Db 595 oGlnGlyProArg-----GlyVa 601  
 QY 2579 GATGGCAACAGAGGACCAACAGCCACATCTCTGATAAAGTAAAGGGGGGTGGATCA 2520  
 Db 601 lGlnGlyProProGlyProThrGlyLysProGlyArgGlyArgAlaGlySerAsp-- 620  
 QY 2519 GCAAAAGACAGTCTGTGGGCTGAGGGACCTGTTCTGTGTGTGTGCTGCTCAGGACT 2460  
 Db 620 ----- 620  
 QY 2459 CTTCCCTACAATAAAGTCATATATGTTCAATCCCATGGAGAGTGTTCATCTAGAAAC 2400  
 Db 620 ----- 620  
 QY 2399 TCCCATGCAAGAGCTACATTAACGAAGCTGCAGGTTAAGGGCTTTAGAGATGGGAACC 2340  
 Db 621 -----GlyAlaArg-----GlyMetPr 626  
 QY 2339 AGTGAAGTATTTATTACAGTCTCCCAAAACCCTTCTCTAGGTGTCTCTCACTAGGAG 2280  
 Db 626 oGlyGln-----ThrGlyProLysGly-----AspArgGlyPheAspGlyLeuAlaG 642  
 QY 2279 CTAGCTGTTAAACCTGAGCTGGGTAAATCCACCTGCAGAGTCCCGCATTTCCAGTGCATG 2220  
 Db 642 Y-----LeuProGlyGluLysGlyHisArgGlyAspPr 653  
 QY 2219 GAGCCCTTCGGCTCCCTGTATATAGTCCAGACTGAAACCCCTTGGAAAGCCTCCAGTC 2160  
 Db 653 oGlyProSerGlyProPro----- 659  
 QY 2159 AGGCAGCCTTAGAGACTGGGAGAGAG-----AGAGGGAC 2124  
 Db 660 -GlyLeuProGlyAspGlyGluArgGlyAspGlyGluValGlyProArgGlyLe 679  
 QY 2123 GCCCCAGCCCCAGCTGTGCAG-----CTACGCACCTCAGCAGCAGAGGTGGCAGCAG 2070  
 Db 679 uProGlyGluProArgGlyLeuLeuGlyProLysGlyProProGlyProGly 699  
 QY 2069 AGAGCCACATTTACTTTGGCAGCAGACAAACTGGGGCCAGCCCGCAGCCCCCATGGGGC 2010  
 Db 699 yProPro-----GlyValThrGlyMetAspGlyGlnProGly---ProLysGlyVa 715  
 QY 2009 TAACAGGA---GCGGGAGCTGGGACCCAGTGGGAGGCCCTCCACCCCAATGTCTGG 1953  
 Db 715 snValGlyProGlnGlyGluProGlyProProGlyGlnGlnGlyAsnProGly----- 732  
 QY 1952 AGTTTTTACCGCTAGTATTTGGCAAGTCGCTCTTGTCATAATACTACTCTGTAGCAA 1893  
 Db 733 -----AlaGlnGlyLeu-ProGlyProGlnGlyAlaIleGlyProProGlyGlu 748  
 QY 1892 AGTAATGGCGACACCCAGCTGGCGCAGACACACATATAGGATGACAGACTGGC 1833  
 Db 749 LysGlyProLeuGlyLysProGlyLeuProGlyMetProGlyAlaAspGlyProProGly 768  
 QY 1832 TGAGCTGGAGATGAGGCCCAATAAACAGGATGGGCCACCTGGG---ACACAGGAAG 1776  
 Db 769 HisProGlyLys-----GluGlyProProGlyGluLysGlyGln 782

QY 1775 CACTATCCAGGATGCGGAGGTCCAGCAGATGCCCGGCCCGGAA----- 1731  
 Db 783 GlyProProGlyProGlnGlyProIleGlyTyrProGlyProArgGlyValLysGlyAla 802  
 QY 1731 ----- 1731  
 Db 803 AspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAspGlyPheProGly 822  
 QY 1730 -----CCACCTGGCCTCGG 1716  
 Db 823 PheLysGlyAspMetGlyIleLysGlyAspArgGlyGluIleGlyProProGlyProArg 842  
 QY 1715 TGGGCTCACCCACACACACACAGTACGGAGACATACACAGCAGAGCCCGCAGACCGCGG 1656  
 Db 843 --GlyGlu-----AspGlyProGluGlyProLysGlyArgG 854  
 QY 1655 GTGAGGTGGGAGCAGCAGCTCCTCCAGCACCCACCTGTCCATTAGGGAAG---GGAG 1599  
 Db 854 LyGlyProAsnGlyAspProGlyProLeuGlyProThrGlyLysGlyLysLeuGlyV 874  
 QY 1598 CTCAGGCTTAGGG-----CCTGGCAGGAAGCTGGTCAATCAGGCTGTCTCACTGCTAG 1545  
 Db 874 aLProGlyLeuProGlyTyrProGlyArgGln-----G 885  
 QY 1544 CACCTCCAGTGTCCCTCGGTATT-----TGGCAGGAACACCTGTCTTCCCGGTGGT 1491  
 Db 885 LyProGlySer-IleGlyPheProGlyPheProGlyAlaAsnGlyGluLysGlyGly 904  
 QY 1490 AGAGGAGGCGAGTGTGTAGGCGAGGATCTGCAGGCTGAGAAAGTGAACCGGTGAGGG 1431  
 Db 905 ArgGlyThrPro-----GlyLysProGly----- 912  
 QY 1430 CGGTGAAGTGTCTACACCGGCCACACTGTGGGACAGCATGTGSCACCGCAGCCACAG 1371  
 Db 913 -----ProArg 914  
 QY 1370 GGAAAGTCCACACTGGCCAAATAGACTGCTCGAGTGCCTGATCGTGCACCGCGGT 1311  
 Db 915 GlyGln-----ArgGlyProThrGly 921  
 QY 1310 CCATGACACAGAGAGA-----AGACCAGG---AGATGGCGCACT 1275  
 Db 922 ProArgGlyGluArgGlyProArgGlyIleThrGlyLysProGlyProLysGlyAsnSer 941  
 QY 1274 GCAGAAACAGCCCA-----GGCTGCCATCCCAAGCCTTCATCATGTCTCGGGCCT 1218  
 Db 942 GlyGlyAspGlyProAlaGlyProProGlyGluArg-----GlyPro 955  
 QY 1217 CGGTCCCGGCTCAGCTCTGGGCGAGCCCTGTACAGCCCTCGCCAGCAAAATCCGTGT 1158  
 Db 956 AsnGlyProGlnGlyProThrGlyPheProGlyProLysGlyProProGlyProGly 975  
 QY 1157 AAAACAGCGTGAAGTTCATAGTGCCA-----LeuProGlyHisProGlyGlnArgGlyGluThrGlyPhe 991  
 Db 976 LysaspGly----- 1000  
 QY 1130 TCCAGCTCAGCTCAGCCACAGAGAGCCCGGCGAGGTCGGGGCATGCGGAGCAGAC 1071  
 Db 992 GlnGlyLysThrGlyProProGlyProProGlyVal---ValGlyProGlnGlyProThr 1010  
 QY 1070 -----GCTGTGCGCGCGGGAAGCAGCGGCGGCTTCCGGAAGCCCAAGCGG--- 1020  
 Db 1011 GlyGluThrGlyProMetGlyGluArgGlyHisProGlyProProGlyProProGlyGlu 1030  
 QY 1019 -----CCCGCATGGACAGCAGTGGGCGCAAGAGGGGGCGCAGACAGCCCTCTCTG 969  
 Db 1031 GlnGlyLeuProGlyValAlaGlyLysGluGlyThrLysGlyAspPro----- 1046  
 QY 968 CTGGCTCGGTGGGCGCCAGCG---CTGCTCTCTCAG-----CCACAGCAGTGTGCTG 918  
 Db 1047 -----GlyProAlaGlyLeuProGlyLysAspGlyProPro----- 1058

917	QY	CTACCCAGGTGAGGAAGATGAGGCTGACGAGGCCCAAGAGGCACAT-	-----	-----	-----	873
1059	Db	-----GlyLeuArgGlyPheProGlyAspArgGlyLeuProGlyProValGly	1074	1074	1074	
872	QY	-----CCTCCTCGGTGCCAGGT-	-----	-----	-----	846
1075	Db	AlaLeuGlyLeuLysGlySerGluGlyProProGlyProProGlyProAlaGlySerPro	1094	1094	1094	
845	QY	GGGCACTGGTGTCCCACTCAATGGCAGGAGGAGGTAGCCCAAGCAGGCCCAAGACTGA	786	786	786	
1095	Db	GlyGluArgGlyProAla-	-----	-----	-----	1109
785	QY	TCATGACGCATAGACACAGTAGTGGCTGCGCAGACTGGTGGCGGTCCCGGAAGAGGTCA	726	726	726	
1110	Db	-----GlyArgProGlyProGlnGlyPro-	-----	-----	-----	1117
725	QY	AGAGCAGGGCCCTCCAGTGAGGTGAAGACACACTGGCCACAGAACTCCAGACGCCCAAGC	666	666	666	
1118	Db	--ProGlyProAlaGlyLysGlyValProGlyGluLysGlyProGlnGlyPro-Al	1136	1136	1136	
665	QY	CCAGATGAGCACTGCCAGCTCCAGGGCCCTGGGATCCGGGCACAGCAGCCCTGCTAGCC	606	606	606	
1136	Db	aglyArgAsp-	-----	-----	-----	1153
605	QY	AGCGCGCCCTTGGGATGAG-	-----	-----	-----	570
1153	Db	oValGlyProProGlyGluAspGlyAspLysGlyGluIleGlyGluProGlyGlnLysG	1173	1173	1173	
569	QY	TGCCCAAGACAGTCCCACTAGATGAAGGGCGGGCGGCCATAGCGTCCAGCCAGTGT	510	510	510	
1173	Db	ySerGlyAspLysGlyGluGlnGlyProGly-ProThrGlyPro--GlnGlyP	1192	1192	1192	
509	QY	CACTGGCTCAGCT--AGGAGCGGCACACAGACCAGCGCCAGCACTGGACCAATGCCCA	453	453	453	
1192	Db	roileGlnProGlyProSerGlyAlaAspGlyGluPro--GlyProArgGlyG	1210	1210	1210	
452	QY	GCACCATGTCTATGACTTCTCTCTACCCCACTTCCAGCAGCAGAGCGGCACATAGG	393	393	393	
1210	Db	lnGlnGlyLeu--PheGlyGlnLysGlyAspGlySerArgGly-	1224	1224	1224	
392	QY	TGATGCTCGCGCCAAACACACTCCAGCC--CAAAGGTTAGCAGGTTGACCAGCAAGA	336	336	336	
1225	Db	--PheProGly-ProProGlyProValGlyLeuGlnGlyLeuProGlyProGlyGlu	1243	1243	1243	
335	QY	GCTGGCTTTCCGGT-	-----	-----	-----	315
1244	Db	LysGlyGluThrGlyAspValGlyGlnMetGlyProProGlyProProGlyProArgGly	1263	1263	1263	
314	QY	GCAGGGCGCTCACCCACAGCCCTCGACCATAGTGGGCCA-	275	275	275	
1264	Db	ProSerGlyAlaProGlyAlaAspGlyPro-GlnGlyProProGlyGlyIleGlyAsnPr	1283	1283	1283	
274	QY	-----GGCGGTTAGGCTCAGGGCGCCCTTCCAGCAGCTCCAGAGACTGTTCTGCT	225	225	225	
1283	Db	oGlyAlaValGlyLysGlyGluProGlyGluAlaGlyGluProGlyLeuProGlyG	1303	1303	1303	
224	QY	CGGC-	-----	-----	-----	207
1303	Db	uGlyGlyProLeuGlyProLysGlyGluArgGlyGluLysGlyGluAlaGlyProSerG	1323	1323	1323	
206	QY	TGCGGGCTCTCCTCCTTGTCTGGCC-	-----	-----	-----	162
1323	Db	yAlaAlaGlyProProGlyProLysGlyProProGlyAspAspGlyProLysGlySerPr	1343	1343	1343	
161	QY	AGGCGCCCATTTCTGCCAGCCCTTTGCT-	-----	-----	-----	114
1343	Db	oGly-	-----	-----	-----	1358
113	QY	CATGCTCAACACCTGCTGTGTGGGCACCTCAGTGGGCACAGCTCTCATCTCAGATC	54	54	54	
1358	Db	-----Gly-ProAlaGlyGlnAspGlyProProGlyAspLysGlyAspAspGlyGluP	1376	1376	1376	
53	QY	CTGGCGGAGCGCGCGCTGTCAACCGGA	25	25	25	

```

Db      1376 roGlycinThrGly-----SerProGly 1383
|||||...|||
RESULT 15
JQ2389
sucrose transport protein - potato
C:Species: Solanum tuberosum (potato)
C:Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000
C:Accession: JQ2389; S40310
P:Riesmeier, J.W.; Hirner, B.; Frommer, W.B.
Plant Cell 5, 1591-1598, 1993
A:Title: Potato sucrose transporter expression in minor veins indicates a role in phloem loading
A:Reference number: JQ2389; MUID:94146554; PMID:8312741
A:Accession: JQ2389
A:Molecule type: mRNA
A:Residues: 1-516 <RIE>
A:Cross-references: EMBL:X69165; NID:g439293; PIDN:CAA48915.1; PID:g439294
A:Experimental source: cv. Desiree
C:Comment: The gene encoding for this protein is highly expressed in mature leaves.
C:Superfamily: common tobacco sucrose transport protein
C:Keywords: glycoprotein; transmembrane protein
F:31-53/Domain: transmembrane #status predicted <TM1>
F:67-86/Domain: transmembrane #status predicted <TM2>
F:103-122/Domain: transmembrane #status predicted <TM3>
F:141-160/Domain: transmembrane #status predicted <TM4>
F:180-200/Domain: transmembrane #status predicted <TM5>
F:226-248/Domain: transmembrane #status predicted <TM6>
F:285-304/Domain: transmembrane #status predicted <TM7>
F:331-349/Domain: transmembrane #status predicted <TM8>
F:366-385/Domain: transmembrane #status predicted <TM9>
F:409-427/Domain: transmembrane #status predicted <TM10>
F:429-448/Domain: transmembrane #status predicted <TM11>
F:3.92/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.:          1,44e-12           Length:         516
Score:              333.00             Matches:        133
Percent Similarity: 41.93%             Conservative:    93
Best Local Similarity: 24.68%           Mismatches:     201
Query Match:        5.19%              Indels:         112
DB:                  2                  Gaps:           21

US-09-759-143-110 (1-3410) x JQ2389 (1-516)

```



```

Db      151 LeuAspValAlaAsnMetLeuGlnGlyProCysArgAlaLeuLeuAlaLeuSer 170
QY      737 CGGACCGGAC--CACTGTCGCGAGGCTACTCTGTGTATGCTTCTATGATCAGTCTT 793
Db      171 GlyGlyLysSerGlyArgMetArgThrAlaAsnAlaPhePheSerPheMetAlaVal 190
QY      794 GGGGGCTGGCTGGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 832
Db      191 GlyAsnIleLeuGlyThrAlaAlaGlySerTyrSerHisLeuPheLysValPheProPhe 210
QY      833 ---GACACACAGTCCCTGGCCCTTACCTACCTGGGACCCAGAGAGAGTGCCTCTTT---GSC 886
Db      211 SerLysThrLysAlaCysAspMetTyrCysAlaAsnLeuLysSerCysPheIleAla 230
QY      887 CTGCTCACCTCATCTTCCTCCTCCTGCTAGCAGCCACACTCTCTGGTGGCTCAGGAG--- 943
Db      231 IlePheIleLeuLeuSerLeuThrIleAlaLeuThrLeuValArgGluAsnGluLeu 250
QY      944 GCAGCGCTGGGCCCCACCGAGCCAGCAGAGGGCTGTGCGGCCCTCTCTGTCGCCCCAC 1003
Db      251 ProGluLysAspGluGlnGluIleAspGluLysLeuAlaGlyAlaGly----- 266
QY      1004 TCTGTCCATGCGGGCGCGCTTGGCTTTCGGAACCTGGGCGCCCTGCTTCCCGCGCTG 1063
Db      267 -----LysSerLysValProPhe-----PheGlyGluIlePheGlyAlaLeu 280
QY      1064 CACCACTGTGCTGCGCGATGCCCGCACCTTCCGCGGCTCTTCTGCTGGCTGAGTGTGC 1123
Db      281 LysGluLeu-----ProArgProMetTrpIleLeuLeuLeuValThrCysLeu 296
QY      1124 AGCTGGATGGCACTCATGACCTTCACGCTGTTTACAGGATTCGTGGCGAGGGCTG 1183
Db      297 AsnTrpIleAlaTrpPhePheLeuTyrAspThrAspTrpMetAlaLysGluVal 316
QY      1184 TACCAGGCGCTGCCAGACTGAGCGGCGCACCGAGCGCGGCGGCGGACATATGATGAGGC 1243
Db      317 PheGlyGly-----GlnValGlyAsp-----AlaArgLeuTyrAspLeuGly 330
QY      1244 GTTCGATGGCGAGCTGGGGCTTCTGTCAGTGGCGCATCTCCCTGGTCTCTCTCTG 1303
Db      331 ValArgAlaGlyAlaMetGlyLeuLeuGlnSerValLeuGlyPheMetSerLeu 350
QY      1304 GTCATGGACCGCTGGTGCGAGCATTCGCGCACTCGAGCAGTCTATTGCGCCAGTGGCA 1363
Db      351 GlyValGluPheLeuGlyLysIleGly-----GlyAlaLysArgLeuTrpCylIleLeu 369
QY      1364 GCTTTCCTGCTGCGCTGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1417
Db      370 AsnPheValLeu-----AlaIleCysLeuAlaMetThrIleLeuValThrLysMet 386
QY      1418 -----ACAGCTTCAGCGCGCCCTCACCGGGCTTCACC----- 1447
Db      387 AlaGluLysSerArgGlnHisAspProAlaGlyThrLeuMetGlyProThrProGlyVal 406
QY      1448 -----TTCACGCCCTGCAGATC----- 1465
Db      407 LysIleGlyAlaLeuLeuPheAlaAlaLeuGlyIleProLeuAlaAlaThrPheSer 426
QY      1466 CTGCGCTACACACTGCGCTCCTCTTACCACGGGAGAG----- 1504
Db      427 IleProPheAlaLeuAlaSerIlePheSerSerAsnArgGlySerGlyGlnGlyLeuSer 446
QY      1505 -----CAGTGTTCTCTGCCCAATATCCAGGGGACACTGGAGGTGCT 1546
Db      447 LeuGlyValLeuAsnLeuAlaIleValProGlnMetLeuValSerLeuValGlyGly 466
QY      1547 ACAGTGGAGCAGCTGATGACACAGCTTCTGCGAGCGCCCTTAAGCCTGGAGCTCCCTTC 1606
Db      467 ProTrpAspLeuPheGlyGlyGlyAsnLeuProGly-----Phe 480
QY      1607 CTAATGGACAGCTGGTGGTGGAGGCACTGGC-----CTGCTCCCACT 1651

```

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

A:Molecule type: DNA  
A:Residues: 1-36 <LEE>  
A:Cross-references: GB:138808; NID:q1020325; PIDN:AAAT9853.1; PID:q1020326  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit are 5-hydroxylated and subsequently O-glycosylated.  
C:Comment: A long form of the mature protein containing part of the amino-terminal propeptide the heterotrimers are probably processed to the long form.  
C:Genetics:  
A:Gene: GDB:COL5A1  
A:Cross-references: GDB:131457; OMIM:120215  
A:Map position: 9q34.2-9q34.3  
C:Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer of alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the length, is formed with desmosine cross-links made from lysine and allysine residues  
C:Function:  
A:Description: structural component of extracellular fibrous polymer associated with cell A:Note: may play a role in controlling the lateral growth of collagen I fibrils  
C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline  
F:1-37/Domain: signal sequence #status predicted <SIG>  
F:36-261/Domain: PARP-like #status predicted <PARP>  
F:38-241/Domain: amino-terminal propeptide #status predicted <PRO>  
F:542-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>  
F:542-558/Region: amino-terminal nonhelical telopeptide  
F:559-1572/Region: helical  
F:645-647/Region: cell attachment (R-G-D) motif  
F:663-665/Region: cell attachment (R-G-D) motif  
F:897-929/Region: heparin binding  
F:1573-1605/Region: carboxyl-terminal nonhelical telopeptide  
F:1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
F:1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
F:38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
F:62-244,183-237/Disulfide bonds: #status predicted  
F:159,176,385,1672,1741/Binding site: carboxydrate (Asn) (covalent) #status predicted  
F:234,236,240,262,263,273,274,275,277,279,280,330,340,346,347,352,357,416,417,420,421/B  
F:335/Modified site: allysine (Lys) #status predicted  
F:341-542/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted  
F:570,576,621,639,648,654,657,675,678,690,693,696,705,717,720,726,732,741,750,753,756,78  
F:627,642,687,708,744,774,795,804,807,810,819,825,846,864,882,897/Modified site: 5-hydro  
F:627,642,687,774,795,804,807,810,819,825,846,864,882,897,1483/Binding site: carboxydrat  
F:708,744/Binding site: carboxydrate (Lys) (covalent) #status experimental  
F:1482/Modified site: 5-hydroxylysine (Lys) #status predicted  
F:1605-1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted  
F:1639,1645,1662,1671/Disulfide bonds: interchain #status predicted  
F:1680-1835,1746-1789/Disulfide bonds: #status predicted

Alignment Scores:  
Pred. No.: 1,44e-12 Length: 1838  
Score: 333.00 Matches: 311  
Percent Similarity: 31.57% Conservative: 59  
Best Local Similarity: 26.54% Mismatches: 381  
Query Match: 5.35% Indels: 424  
DB: 1 Gaps: 68

US-09-759-143-110 (1-3410) x CGHUV1 (1-1838)

QY 3217 AAACGGCACTTAACCCCTCCGAGATGAAGACCTCCCTTACGCTCAG----- 3170  
Db 708 LysGlyAsnValcylProGlnGlyGluProGlyProGlyGlnGlnGlyAsnProGly 727  
QY 3169 GCAGGGGTGCTCCTCAGTTCTGTGTGAGATCCCGAAGCAGATATCTCTGGGGC 3110  
Db 728 AlaGlnGlyLeuProGly-----ProGlnGlyAlaIleGlyProProGly 742  
QY 3109 TGAGATGGACAAGGCTTGGGAACCGCAGCTTTGT-----GCTTCTGCTCT 3063  
Db 743 GluLysGlyPro---LeuGlyLysProGlyLeuProGlyMetProGlyAlaAspGlyPro 761  
QY 3062 GCAGTAGCTCCAAACAGGTTGTGGAGCTGGTGGGGAAGTTGGGGGTAGGGGAAGTTG 3003  
Db 762 ProGlyHisProGlyLysGlu--GlyProProGlyGluLysGlyGlnGlyProProG 781

3002 GGGTAGGGGAARATTTTGGCCAGTGCCTTCATCAGCCAGTCTCTAGACAGAGTAGAGGG- 2944  
Db 781 LysProGlnGlyProLleGly-----TyrProGlyProArgGlyValLysGlyA 797  
QY 2943 -----GAGTGGAAAGTGGGGGAACAC 2922  
Db 797 laAspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAspGlyPheProG 817  
QY 2921 GCTGGGCCAAGAGAAGAGGGGTGGTTAGGGAAGCCCTTACAGACCTGAAGCCGCCCTCT 2862  
Db 817 LysPheLysGlyAspMetGlyIleLysGlyAspArgGlyGluIleGlyProPro----- 834  
QY 2861 ACCTTCTTCAACACCCTAACCTTGGGTAAACAGCATTTGGAATTTATCATTTGGGATGAGT 2802  
Db 834 ----- 834  
QY 2801 AGAATTTCCAAAGTCTCTGGTGGTGGTGGGAGCATTTTGGGGGCCAGACCCCGAGAGAAAGATT 2742  
Db 835 -----ArgGlyGluAspGlyProGluGlyProL 846  
QY 2741 CTGGCAATGATCAGCCCAATGA-----CCAGCTATCTCAGGGG 2704  
Db 846 ysGlyArgGlyGlyProAsnGlyAspProGlyProLeuGlyProGlyGluLysGlyL 866  
QY 2703 ACCTGATTGTTGGGATCCCGCCACCTACCCAATATTAGACACCAACAGAAAAGCTA 2644  
Db 866 ysLeuGlyValProGlyLeuProGlyTyrProGly-----ArgGlnGlyProLysGlyS 884  
QY 2643 GCAATGATGATCCCTTCTACTTTTCTTAATAAATAAGTTAAATATTTAAATGCTGTGTCT 2584  
Db 884 erlleGlyPheProGly-----PheProGlyAlaAsnGlyGluLysG 898  
QY 2583 CTGTGATGCAACAGAAAGACCAAGCCACATCTGTATAAAGTAAGTAGAGGGGGTGG 2524  
Db 898 LysGlyArgGlyThrProGlyProGlyPro-----ArgGlyGlnArgGly--- 913  
QY 2523 ATCAGCAAAAGACAGTGTGTGGCTGAGGGGACCTGTTCTTGTGTGTGTGGTCCCTCAG 2464  
Db 914 -----ProThrG 916  
QY 2463 GACTCTTCCCTACAAATAAGTCATATGTTCAATCCATGGAGAGTGTTCATCCTAG 2404  
Db 916 Lys----- 916  
QY 2403 AAATCCCATGCAAGACCTACATTAACGAAGTGCAGGTAAAGGGCTTAGA-----G 2350  
Db 917 -----ProArgGlyGlu-----ArgGlyProArgGlyL 927  
QY 2349 ATGGGAACCAAGGTGACTGAGTTTATTTCAGCTCCCAAAACCCTTCTCTAGTGTGTCTC 2290  
Db 927 hrGlyLysProGly-----ProLysGlyAsnSerGlyGlyAspGlyP 941  
QY 2289 AACTAGGAGGCTAGCTGTTAAACCTGAGCCTGGGTAAATCCACCTGCAGAGTCCCGCAT 2230  
Db 941 roAlaGly-----ProProGlyGluArgGlyProA 951  
QY 2229 CCAGTGTGATGGAGCCCTTCTGGCCCTCCTGTATAGTCCAGACTGAACCCCTC-----T 2176  
Db 951 snGlyProGlnGlyProThrGlyPheProGlyProLysGlyProGlyProProGlyL 971  
QY 2175 TGAAGGCTCCAGCTCAGGAGCCCTAGAGACTGGGAGAG-----AGGA 2131  
Db 971 ysAspGlyLeu---ProGlyHisProGlyGlnArgGlyGluThrGlyPheGlnGlyLys 990  
QY 2130 GAGGAGCCGCCCGCCCGCTGTGCAGCTGTACGCTACCCACTCAGCAGCAGAGGTGGACGA 2071  
Db 990 hrGlyProProGlyProProGlyVal-----ValGlyProGlnGlyProThrG 1006  
QY 2070 GAGAGCCCATTTACTTTGGCAGCAACAACTGCAGCCAGCCCGCCAGCC----- 2017  
Db 1006 LysGluThrGlyProMetGlyGluArgGlyHis--ProGlyProProGlyProProGlyGlu 1025

```
QY 2016 ATGGGGCTAACAGAGCGGGAGCTGGGACCCAGTCAGGAGCGCCCTCCACCCCAATGTG 1957
Db      ||||| ||| ||||| ||| ||| ||||| |||||
QY 1026 GlnGlyLeuProGlyLeu-Ala- 1037
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1956 CTGGAAGTTTCTACGCTGAGTATTGGCCAAAGTCCTCTGTGCAAAATACTACCTGTGTA 1897
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1037 1037
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1896 GCAAGTAATGGCAGCAGACCCAGGCTGGGGCAGACACCATATAGGCAGTCAGAC 1837
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1038 1038
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1836 TGCGCTGAGCTGGAACAATGGAGCCCAATAACAGGGATGGGGCCACCTGGGACAGCAGGAAG 1777
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1051 1051
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1776 GCACATCCAGGATGGCGAGGTCAGGCGAGATGCCCGGC 1737
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1058 1058
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1736 1736
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1073 uLysGlyAsnGluGlyProProGlyProAlaGlyLeuProGlyProAlaGlyLeuGlyLe 1073
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1695 CGTACGAGACATCAGCAGCAGGCGCGCAGAGCGGGGTGGAGTGGGAGCAGGCCA 1636
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1090 1090
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1635 CTGCTCCAGCAGCCAGCTGTCCATTAGG 1591
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1108 GlyProGlnGlyProProGlyProAlaGlyGluGlyGlyAlaProGlyGluGlyLysGlyPro 1127
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1590 TTAGGCGCTGGCAGGAAGCTGTGCATCAGGCTGTCTCCCGGTGGTAGAGGAGCCAGCTGTGTAG 1471
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1128 GlnGlyPro-AlaGly-ArgAspGly- 1139
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1530 CCTCGGTATTGGGAGGAGACACCTGCTTCTCCCGGTGGTAGAGGAGCCAGCTGTGTGTAG 1471
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1139 oValGlyLeu- 1154
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1470 GGCAGGATCTGAGGCTGAGAAAGCTGAACCGGTGAGGGGGTGAAGCTGTACACG 1411
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1154 uAspGlyAsp-LysGlyGluIleGlyGluProGlyGlnIleGlySerLysGlyAspLysG 1174
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1410 GCCAC- 1384
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1174 LyGluGlnGlyProGlyProThrGlyProGlnGlyProIleGlyGlnProGlyProS 1194
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1383 CCGGAGCAGCACAGGAAAGCTGCCACACTGGCCCAATAGACTGCTCGAGTCCCAATGCG 1324
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1194 erGlyAlaAspGlyGluProGlyProArgGlyGln- 1205
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1323 TGCCACCGCGGTCCATCAGCAGAGAGACACAGGAGATGGC 1280
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1206 1206
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1279 1279
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1224 roGlyProValGlyLeuGln-GlyLeuProGlyProProGlyGluLysGlyGluThrG 1243
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1230 TGTCTCCGGGCTGGTCCCGGCTCAGCTCTGGGCGAGCCCTGTACAGCCCTCGCC 1171
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1243 yAspValGlyGlnMetGlyPro- 1256
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1170 AGCAATCCGTGTAAACAGCGTGAAGTGCATGATGCCATCCAGCTGCACAGCTCAGCC 1111
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1256 oArgGlyPro- 1270
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1110 ACAGAGACCGCGCAGG- 1060
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1270 yProProGlyGlyIleGlyAsnProGlyAlaValGlyLysGlyGluProGlyGluAl 1290
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1059 CGGGGAAGCAGGCGCCCGCTTCGGGAAACCAAGCGGGCGCGCATGGACAGCTGG 1000
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1561 yProProGly 1564

RESULT 17

S23810

collagen alpha 1(XVI) chain precursor - human

N:Alternate names: procollagen alpha 1(XVI) chain

C:Species: Homo sapiens (man)

C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 20-Sep-1999

C:Accession: S23810; PQ0612; S08012

R:Pan, T.C.; Zhang, R.Z.; Mattel, M.G.; Timpl, R.; Chu, M.L.

Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992

A:Title: Cloning and chromosomal location of human alpha1(XVI) collagen.

A:Reference number: S23810; MUID:92335339; PMID:1631157

A:Accession: S23810

A:Molecule type: mRNA

A:Residues: 1-1603 <PAN>

A:Cross-references: EMBL:M92642; NID:g180757; PIDN:AAA58427.1; PID:g180758

A:Experimental source: skin fibroblasts

R:Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J. Biochem. 112, 856-863, 1992

A:Title: Molecular cloning and partial characterization of a novel collagen chain, alpha 1(XVI)

A:Reference number: PQ0612; MUID:93203161; PMID:1284248

A:Accession: PQ0612

A:Molecule type: mRNA

A:Residues: 'GGR', 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1164, 'P', 1166-1603 <YAM>

A:Cross-references: GB:S57132; NID:g298641; PIDN:AAB25797.1; PID:g298642

A:Experimental source: placenta

R:Kimura, S.

submitted to the EMBL Data Library, April 1989

A:Description: Partial nucleotide and amino acid sequence of a collagen-like protein from placenta

A:Reference number: S08012

A:Accession: S08012

A:Molecule type: mRNA

A:Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM' <KIM>

A:Cross-references: EMBL:X14963; NID:g29984; PIDN:CAA33085.1; PID:g930048

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit are modified and subsequently O-glycosylated.

C:Genetics:

A:Gene: GDB:COL16A1

A:Cross-references: GDB:134045; OMIM:120326

A:Map position: lp34-lp34

C:Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XVI)

C:Function:

A:Description: structural component of extracellular fibrous polymer as a minor form procollagen

A:Note: may play a role in forming elastic connections at fibril surfaces

C:Superfamily: unassigned collagens

C:Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxylysine

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAT>

F:22-333/Domain: amino-terminal nonhelical #status predicted <NC1>

F:334-1577/Region: interrupted helical

F:334-360/Domain: collagenous COL10 #status predicted <COL0>

F:375-505/Domain: collagenous COL9 #status predicted <COL9>

F:521-554/Domain: collagenous COL8 #status predicted <COL8>

F:539-541/Region: cell attachment (R-G-D) motif

F:572-630/Domain: collagenous COL7 #status predicted <COL7>

F:652-722/Domain: collagenous COL6 #status predicted <COL6>

F:738-875/Domain: collagenous COL5 #status predicted <COL5>

F:887-938/Domain: collagenous COL4 #status predicted <COL4>

F:973-987/Domain: collagenous COL3 #status predicted <COL3>

F:1005-1007/Region: cell attachment (R-G-D) motif

F:1011-1432/Domain: collagenous COL2 #status predicted <COL2>

F:1226-1228/Region: cell attachment (R-G-D) motif

F:1472-1577/Domain: collagenous COL1 #status predicted <COL1>

F:1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NC01>

F:47,327/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Alignment Scores:

Pred. No.: 2,18e-12 Length: 1603

Score: 330.00 Matches: 316

Percent Similarity: 31.43% Conservative: 81

Best Local Similarity: 25.02% Mismatches: 417

Query Match: 5.30% Indels: 451

DB: 2 Gaps: 69

US-09-759-143-110 (1-3410) x S23810 (1-1603)

Qy 3077 TGTGCTTCTGTCCT-----GCAGTAGCTCCAAACAGGGTT 3042

Db 372 CysAlaGluGlyProLysGlyGluLysGlyGluSerGlyAlaLeuGlyProSerGlyLeu 391

Qy 3041 GTGGAGCTGTGGGAAAGTTGGGGTAGGGGAGGAGTTGGGGAGGAAATTTTGGG 2982

Db 392 ProGlySerThrGlyGluLysGlyGlnLysGlyGluLysGlyAspGlyGlyLeuLysGly 411

Qy 2981 AGTGCCTTCATCAGCCCGAGTCTCTAGAGAGAGTACAGGGAG----- 2941

Db 412 -ValProGly-LysProGly---ArgAspAlaProGlyGluLeuValLeuGlyProL 430

Qy 2940 --TGAAGTGGGGGAACACAGGCTGGCCCAAGAGAGAGGGTGTAGGAAAGCCGTTG 2883

Db 430 ysGlyGlnLysGlyAspProGlyPheValGlyProGluGlyLeuAlaGlyGluPro--- 448

Qy 2882 AGACCTGAAGCCCGCCCTCTACCTCTCTTCAACACCCCTAACCTTGGGTAAACAGCATTTG 2823

Db 449 -----GlyPro-----ProGlyLeuProGlyPro----- 456

Qy 2822 GAATTATCATTTGGGTAGTAGTAGAATTCCAGAGTCTCTGGTGTAGGCATTTTGGGGGCC 2763

Db 457 -----ProGlyIleGlyLeuProGly----- 463

Qy 2762 AGACCCGAGGAGAAGATTCTGGCAATGATCAGCCCAATGA-----CCAGCTATCTCAG 2707

Db 464 --ThrProGlyAsp-----ProGlyGlyProProGlyProLysG 476

Qy 2706 GGGACCTGATTGTTGGGATCCCCACCCCTACCCCAATATTAGACACACACAGAAAG 2647

Db 476 Lysp-----LysG 479

Qy 2646 CTAGCAATGGATCCCTCTCTACTTGTAAATAAATAAGTTAAATATTAAATGCCTGTG 2587

Db 479 LysSerGlyIlePro----- 484

Qy 2586 TCTCTGTGATGGCAACAGAGACACAGCCACATCTCTGATAAAGTTAAGAGGGGG 2527

Db 485 -----GlyLysGluGlyProGlyGlyLys-----ProGlyLysProGly 498

Qy 2526 TGGATCAGCAAAAGACAGTCTCTGGCTCAGGGGACCTGGTCTTGTGTGGTCCCT 2467

Db 498 alLysGlyGluLys-----GlyAsp----- 504

Qy 2466 CAGGACTCTCCCTTACAAATAAGTCATATGTTCAAAATCCCATGGAGAGTGTTCATCC 2407

Db 505 -----ProCysGluValCysProThrL 512

Qy 2406 TAGAAATCCCATGCAAGAGCTACATTAAACGAGCTCCAGTTAAGGGCTTAGAGATG 2347

Db 512 euProGluGlyPheGlnAsnPhe-----ValGlyLeuProGlyLysProGlyProLysG 530

Qy 2346 GGAACACAGGTCACTGAGTTTATCCA----- 2320

Db 530 LysGluProGlyAspProValArgAlaArgGlyAspProGlyIleGlnGlyLysGlyG 550

Qy 2319 CTCCTCAAAACCTTCTCTAGTGTGTCTCAACTAGGAGGTAGCTGTATACCTGAGC- 2261

Db 550 LuLysGlyGluProCysLeuSerCysSerValValGlyAlaGlnHisLeuValSers 570

Qy 2260 -----CTGGGTATCCA----- 2249

Db 570 erThrGlyAlaSerGlyAspValGlySerProGlyPheGlyLeuProGlyLeuProGlyA 590

Qy 2248 -----CCTGCAG 2242

Db 590 rgAlaGlyValProGlyLeuLysGlyGluLysGlyAsnPheGlyGluAlaGlyProAlaG 610

Qy 2241 AGTCCCG---CATTCAGTCATGAGCCCTTCTGGCCCTC----- 2204











QY	1818	GAGCCATAAACAGGATGGGGCCA-----CCTGGGACAGCAGGAAG-----	1776
DB	500	aGlyProGlnGlyLysValGlyProSerGlyAlaProGlyGluAspGlyArgProGlyPr	520
QY	1775	-----CACTATCCAGATGCGCAGG	1756
DB	520	oProGlyProGlnGlyAlaArgGlyGlnProGlyValMetGlyPheProGlyProLysG	540
QY	1755	TCAGGCAGATCCCCGGC-----CGGAACACCCCTGCCTCCG	1717
DB	540	YalaAsnGlyGluProGlyLysAlaGlyGlnGlyLysGlyLeuPro-GlyAlaProGlyLeuA	560
QY	1716	GTGGGCTACCCACCAACACACGCTACGAGGACATCACAGGAGAGCCCGCAGAGCCG	1657
DB	560	rgGlyLeu-ProGlyLysasp--GlyGluThrGlyAlaAlaGlyProGlyProAla	578
QY	1656	GTGGAGGTGGAGCAGG-----	1639
DB	579	GlyProAlaGlyGluArgGlyGluGlnGlyAlaProGlyProSerGlyPheGlnGlyLeu	598
QY	1638	CCACTCCCTCCAGACCCACGCTGTCCATTAGGGAAG-----	1603
DB	599	ProGlyProProGlyProProGlyGluGlyLysProGlyAspGlnGlyValProGly	618
QY	1602	-----GGAGCTCCAGGCTTA--GGCCCTGGCAGGAAG-----TGTGTCATC	1564
DB	619	GluAlaGlyAlaProGlyLeuValGlyPro-ArgGlyGluArgGlyPheProGlyGluAr	638
QY	1563	AGGCTGTCTCACTGCTACCACTCCAGGTCCTCCGCTATTGGGACAGGAACACCTGC	1504
DB	638	gGlySerProGlyAlaGlnGlyLeuGlnGlyAlaArgGlyLeu--ProGlyThrProG	657
QY	1503	TTCTCCCGGTGTAGAGGAGG-----CCAGTGTGTAGGCGAGGATCTGC	1459
DB	657	yThrAspGlyProLysGlyAlaSerGlyProGlyProGlyAlaGlnGlyProPr	677
QY	1458	AGGCTGTAGAGTGAACCCGTGAGGGCGCTGAAGCTG-----TCACCAAGCGCCAC	1406
DB	677	oGlyLeu-GlnGlyMetProGlyGluArgGlyAlaAlaGlyLeaAlaGlyProLysGlyA	697
QY	1405	ACTGTGGGACGATGTGSCACCGCAGCGAGGAAAGCTGCCACACTGCCCAATA	1346
DB	697	spArgGly---AspValGlyGluLysGlyProGlyGlyAlaProGlyLysAspGlyGlyA	716
QY	1345	GACTGTCTGAGTCCGAATCCGT-----GCACCAAGCGGTCCATCCACAGAGA-	1296
DB	716	rgGlyLeuThrGlyProIleGlyProGlyProAlaGlyAlaAsnGlyGluLysGlyG	736
QY	1295	-----AGACAGGAGATGGCGACTCGAGGAACAGCCCGAGGCTGCCATCCGAAC	1244
DB	736	luValGlyProProGlyProAlaGlyThrAlaGly--AlaArgGlyAlaProGlyGluA	755
QY	1243	GCCTTTCATCATGTCTCGGGCTCGGTGCGCGCTCACTGCTGGGACGCGCTTGGTA	1184
DB	755	rg-----GlyGluThrGlyProGlyProAlaGlyPheAlaGlyProProGlyA	772
QY	1183	CA-----	1182
DB	772	laAspGlyGlnProGlyAlaLysGlyGlnGlyGluAlaGlyGlnLysGlyAspAlaG	792
QY	1181	--GCCCTCGCCACGAAATCCGTGTAACACAGCGTGAAGTCCATGATGCCATCCAGCT	1124
DB	792	lyAlaProGlyProGlnGlyPro--SerGlyAlaProGlyProGlnGlyProThrGlyV	811
QY	1123	GCACAGCTCAGCCAGGAAGACCCGGCGAGGTCTCGGGGATCGGCGAGCAGACAGCT	1068
DB	811	alThrGly--ProLysGlyAlaArgGlyAlaGlnGlyProProGlyAlaThrGlyPhep	830
QY	1067	--GTGACGCGGGGAAGAGCGCGCCAGGTTC-----GGAAAGCCAAAGCGCGCGCA	1013
DB	830	roGlyAlaAlaGlyArgValGlyProProGlySerAsnGlyAsnProGlyProProGlyP	850

QY	1012	TG-----	-----CAGCAGCTGGCGCA	995
Db	850	roProGlyProSerGlyLysAspGlyProLysGlyAlaArgGlyAspSerGlyProProG	870	
QY	994	CAAGGAGGGGCGCAGAGCCCTTCCTGCTCGGTGGGGCCAGCGCTGCTCCT	939	
Db	870	lyArgAlaGlyAspProGlyLeu	886	
QY	938	-----	-----CAGCCACCAGCAGTGTGGC	920
Db	886	luLysGlyGluProGlyAspAspGlyProSerGlyProAspGlyProPro	902	
QY	919	TGCTACCCAGGTTCAGAAATCAGGCTGAGCGCCAAAGAGGCCTCCTCCTGGGTGCC	860	
Db	903	-----	-----ValGlyLeuP	917
QY	859	CAGGT-----	AGGGG---CCAGGGCACTGTGTCCCACTCAATGGCAGGCAG	815
Db	917	roGlyGlnArgGlyGluArgGlyPheProGlyLeuProGlyProSer	932	
QY	814	GAGGTAGCCAGCCACCCCAAGACTGATCATGAAGCATAGACAGTAGTAGGCTGTGCG	755	
Db	933	--GlyGluProGly	-----LysGlnGlyAlaProGlyA	943
QY	754	ACATGTGT-----	CGGGT-----	719
Db	943	laSerGlyAspArgGlyProProGlyProValGlyProProGlyLeuThrGlyProAlaG	963	
QY	718	GGCTTCAGTGGAGTCAAGCACACTGGCCACAGAGTCCACAGCCGCCAGCCACAGA-	660	
Db	963	lyGluProGlyArgGluGlyThrProGlyAlaAspGlyPro	979	
QY	659	-----	-----TGAGCAGTGTCCAGAGTCCAG	641
Db	979	rgAspGlyAlaAlaGlyValLysGlyAspArgGlyGluAlaGlyAlaLeuGlyAlaProG	999	
QY	640	GGGCC-----	-----TGGGATCCGGGCMACAGCAGCCCTGCTAGCAGCCGCGCTTGG	593
Db	999	lyAlaProGlyProProGlySerProGlyProAlaGly	1014	
QY	592	GA-----	-----TGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGTCCAGATGAAGGGCCG	539
Db	1014	lyLysGlnGlyAspArgGlyGluAlaGlyAlaGlnGlyProMetGlyProAlaGlyPro	1033	
QY	538	GCGCC-----	-----CGCCATAGCGCTC	522
Db	1034	AlaGlyAlaArgGlyLeuProGlyProGlnGlyProArgGlyAspLysGlyGluAlaGly	1053	
QY	521	CACCCAGTGTCTACTGGCTTAGCCTTAGGAGCGGGACACAGA	480	
Db	1054	GluAlaGlyLysGlyLeuLys	-----GlyHisArgGlyPheThrGlyLeuGln	1070
QY	479	-----	-----CGAGCCCACTCGACCA-----	447
Db	1071	GlyLeuProGlyProProGlyProSerGlyAspGlnGlyAlaSerGlyProAlaGlyPro	1090	
QY	446	TGGTCATGAACCTTCCTCTACCCCACTT-----	CCAGCAGCAGAGGGCGCACATAGTCA	390
Db	1091	SerGlyProArgGlyProProGlyProValGlyProSerGlyLysAspGlyAlaAsnGly	1110	
QY	389	TGCCTGCGCCAAACACACCTCCAGGCCAAAGGTTAGCAGTTTGACACAGCAGAGCTGG	330	
Db	1111	lleProGlyProlleGlyProProGlyProArg	-----GlyArgSerGly	1125
QY	329	CTTTCGGT-----	GGCGCAGCAGGGCGTCTACCCACAGCCTCTGACCAATAGTGGCCAGG	273
Db	1126	GluThrGlyProAlaGlyProProGlyYasnProGlyProProGlyProProGlyProPr	1145	
QY	272	CGGG-----	-----TAGGGCTCAGGGCGGCTTCAGGCCTCCAG	237
Db	1145	oGlyProGlylleAspMetSerAlaPheAlaGlyLeuGlyProArgGluLysGlyProAs	1165	
QY	236	ACTGCTTGTCTCGGCTCTGCTCCAGAGCTGCGGCCTCTCCTCCTTGTGTCGGCCNACT	177	